Chromosomal instability drives metastasis through a cytosolic DNA response

Bakhoum, Samuel F ; Ngo, Bryan ; Laughney, Ashley M ; Cavallo, Julie-Ann ; Murphy, Charles J ; Ly, Peter ; Shah, Pragya ; Sriram, Roshan K ; Watkins, Thomas B K ; Taunk, Neil K ; Duran, Mercedes ; Pauli, Chantal ; et al

Abstract: Chromosomal instability is a hallmark of cancer that results from ongoing errors in chromosome segregation during mitosis. Although chromosomal instability is a major driver of tumour evolution, its role in metastasis has not been established. Here we show that chromosomal instability promotes metastasis by sustaining a tumour cell-autonomous response to cytosolic DNA. Errors in chromosome segregation create a preponderance of micronuclei whose rupture spills genomic DNA into the cytosol. This leads to the activation of the cGAS-STING (cyclic GMP-AMP synthase-stimulator of interferon genes) cytosolic DNA-sensing pathway and downstream noncanonical NF- B signalling. Genetic suppression of chromosomal instability markedly delays metastasis even in highly aneuploid tumour models, whereas continuous chromosome segregation errors promote cellular invasion and metastasis in a STING-dependent manner. By subverting lethal epithelial responses to cytosolic DNA, chromosomally unstable tumour cells co-opt chronic activation of innate immune pathways to spread to distant organs.

DOI: https://doi.org/10.1038/nature25432

ARTICLE

doi:10.1038/nature25432

Chromosomal instability drives metastasis through a cytosolic DNA response

Samuel F. Bakhoum[1,2]*, Bryan Ngo[2]*, Ashley M. Laughney[3], Julie-Ann Cavallo[1,2], Charles J. Murphy[2], Peter Ly[4], Pragya Shah[5], Roshan K. Sriram[2], Thomas B. K. Watkins[6], Neil K. Taunk[1], Mercedes Duran[1,2], Chantal Pauli[7], Christine Shaw[8], Kalyani Chadalavada[8], Vinagolu K. Rajasekhar[9], Giulio Genovese[10], Subramanian Venkatesan[11], Nicolai J. Birkbak[6,11], Nicholas McGranahan[6,11], Mark Lundquist[2], Quincey LaPlant[1], John H. Healey[9], Olivier Elemento[2], Christine H. Chung[12], Nancy Y. Lee[1], Marcin Imielenski[2], Gouri Nanjangud[8], Dana Pe'er[13], Don W. Cleveland[4], Simon N. Powell[1], Jan Lammerding[5], Charles Swanton[6,11] & Lewis C. Cantley[2]

Chromosomal instability is a hallmark of cancer that results from ongoing errors in chromosome segregation during mitosis. Although chromosomal instability is a major driver of tumour evolution, its role in metastasis has not been established. Here we show that chromosomal instability promotes metastasis by sustaining a tumour cell–autonomous response to cytosolic DNA. Errors in chromosome segregation create a preponderance of micronuclei whose rupture spills genomic DNA into the cytosol. This leads to the activation of the cGAS–STING (cyclic GMP–AMP synthase–stimulator of interferon genes) cytosolic DNA–sensing pathway and downstream noncanonical NF-κB signalling. Genetic suppression of chromosomal instability markedly delays metastasis even in highly aneuploid tumour models, whereas continuous chromosome segregation errors promote cellular invasion and metastasis in a STING-dependent manner. By subverting lethal epithelial responses to cytosolic DNA, chromosomally unstable tumour cells co-opt chronic activation of innate immune pathways to spread to distant organs.

Chromosomal instability (CIN) correlates with tumour metastasis[1,2], but it remains unclear whether it is a mere bystander or a driver of metastatic progression. Chromosomally unstable cells show evidence of chromosome missegregation during anaphase[3,4], offering an attractive bottleneck in which to target CIN and probe its selective contribution to metastasis. Destabilization of microtubule attachments to chromosomes at the kinetochores, through overexpression of the non-motile microtubule-depolymerizing kinesin-13 family proteins KIF2B or KIF2C (also known as MCAK), directly suppresses CIN in otherwise chromosomally unstable cells[5–7]. Cells overexpressing KIF2B or MCAK continue to propagate abnormal aneuploid karyotypes, albeit in a stable manner[7]. As such, this approach permits direct experimental interrogation of CIN, as defined by the rate of ongoing chromosome missegregation, independently of aneuploidy, which is defined as a state of abnormal chromosome numbers.

Increased CIN in human metastases

First, to determine whether CIN is associated with human metastases, we applied the weighted-genomic integrity index (wGII) as a proxy for CIN[8] to 79 matched pairs of primary tumour and brain metastasis from a published cohort[9]. Metastases showed higher wGII than primary tumours (Fig. 1a and Extended Data Fig. 1a, b).

Next, karyotype analysis of primary breast tumours and metastases archived in the Mitelman Database of chromosomal translocations[10] revealed a predilection for near-diploid (2*n*) karyotypes in primary tumours. Conversely, metastases were enriched for cells with near-triploid (3*n*) karyotypes and had twice as many structural or numerical chromosomal aberrations per clone as primary tumours. The number of chromosomal aberrations was highest in tumour samples with karyotypes ranging between the diploid and tetraploid (4*n*) range (Fig. 1b, c and Extended Data Fig. 1c, d).

Finally, histological analysis of primary tumours from patients with locally advanced head and neck squamous cell carcinoma[11] revealed a significant association ($P < 0.05$) between anaphase chromosome missegregation and the incidence of lymph node metastasis (Fig. 1d and Extended Data Fig. 1e).

CIN is a driver of metastasis

To determine whether CIN is causally involved in metastasis, we used transplantable metastatic tumour models of human (MDA-MB-231) or mouse (4T1) triple-negative breast cancer and human lung adenocarcinoma (H2030), in which 47%, 55%, and 67% of anaphase cells, respectively, show evidence of chromosome missegregation. Overexpression of KIF2B or MCAK suppressed chromosome missegregation, whereas overexpression of a dominant-negative MCAK mutant[12] (dnMCAK) led to a modest increase in chromosome missegregation in MDA-MB-231 cells. Overexpression of KIF2B or MCAK did not alter cellular proliferation or the number of centrosomes per cell (Fig. 2a, b and Extended Data Fig. 1f–i). As a control, we overexpressed KIF2A, a third member of the kinesin-13 family that lacks kinetochore and centromere localization domains[13]; although KIF2A showed microtubule-depolymerizing activity on interphase microtubules, it had no observable effect on CIN (Fig. 2b and Extended Data Fig. 1i–k). We ruled out a direct role for kinesin-13-mediated microtubule depolymerization in

[1]Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [2]Sandra and Edward Meyer Cancer Center, Weill Cornell Medicine, New York, New York 10065, USA. [3]Cancer Biology and Genetics Program, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [4]Ludwig Institute for Cancer Research, University of California San Diego, La Jolla, California 92093, USA. [5]Nancy E. and Peter C. Meinig School of Biomedical Engineering & Weill Institute for Cell and Molecular Biology, Cornell University, Ithaca, New York 14850, USA. [6]The Francis Crick Institute, London NW1 1AT, UK. [7]Institute for Pathology and Molecular Pathology, University Hospital Zurich, Zurich 8091, Switzerland. [8]Molecular Cytogenetics Core, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [9]Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [10]The Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. [11]UCL Cancer Institute, London WC1E 6BT, UK. [12]Moffitt Cancer Center, Tampa, Florida 33612, USA. [13]Computational Biology Program, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA.
*These authors contributed equally to this work.

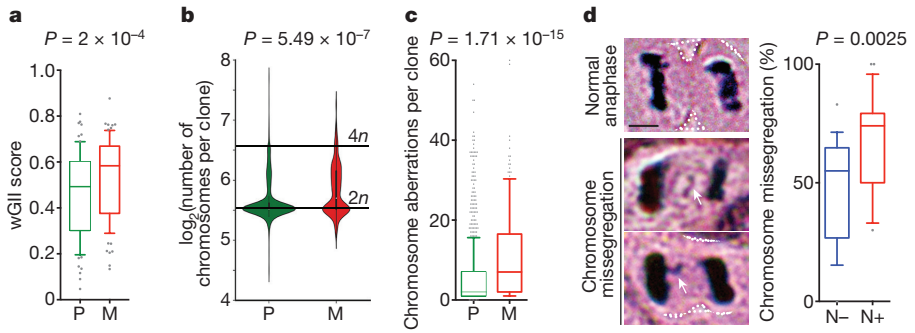


Figure 1 | Human metastases enrich for CIN. a, wGII of matched primary tumours (P) and brain metastases (M), $n = 79$ patients. **b**, **c**, Karyotype probability density (**b**) and chromosomal aberrations (**c**) in 983 primary tumour and 186 metastatic breast cancer clones. **d**, Left, images of head and neck squamous cell carcinoma cells undergoing anaphase. Arrows point to chromosome missegregation; scale bar, 5 μm. Right, chromosome missegregation in tumours from patients with (N+, $n = 22$ patients) or without (N−, $n = 18$ patients) clinically detectable lymph node metastases. Boxes represent median ± interquartile range, confidence intervals (whiskers) denote 10th–90th percentiles (**a**, **c**, **d**), significance tested using two-sided Wilcoxon matched-pairs signed rank test (**a**) and two-sided Mann–Whitney test (**b**–**d**).

activating small GTPases[14] by performing RhoA and Rac1 pull-down assays, which revealed low basal levels of activity and no correlation with overexpression of kinesin-13 family proteins (Extended Data Fig. 2a, b). Hereafter, we refer to cells expressing MCAK or KIF2B as CIN-low and to control cells or those expressing KIF2A or dnMCAK as CIN-high.

Karyotyping of the parental MDA-MB-231 cells revealed a widely aneuploid (approximately 3*n*) chromosome content with widespread karyotypic heterogeneity (Extended Data Fig. 2c). Suppression of CIN reduced both numerical and structural karyotypic heterogeneity in single-cell-derived clones, as supported by the presence of fewer chromosomes exhibiting non-clonal structural abnormalities and decreased numerical chromosome heterogeneity in CIN-low cells (Extended Data Fig. 2c–h). Notably, CIN-low cells maintained highly aneuploid karyotypes, but faithfully propagated them in a stable manner. Thus, by comparing chromosomally stable aneuploid cells to their chromosomally unstable aneuploid counterparts, we can experimentally examine the role of CIN, independent of aneuploidy, in metastasis.

We injected MDA-MB-231 cells into the left cardiac ventricles of athymic mice to enable systemic dissemination while tracking metastatic colonization using a bioluminescence reporter. Differences in chromosome missegregation rates had a marked effect on colonization: mice harbouring CIN-high cells rapidly succumbed to metastatic disease, with a median survival of 70 days, whereas mice injected with CIN-low cells had a lower metastatic burden and a median survival of 207 days. Many metastases from CIN-low cells waxed and waned and, at times, spontaneously resolved, whereas metastases from CIN-high cells involved multiple organs and progressed rapidly, leading to death. Similar results were obtained after injection of lung adenocarcinoma H2030 cells (Fig. 2c–e and Extended Data Fig. 3a–c). Overexpression of the spindle assembly checkpoint protein MAD2 in MCAK-expressing cells partially rescued chromosome missegregation[15] and correspondingly augmented metastasis (Fig. 2c and Extended Data Fig. 3f).

We performed orthotopic injections of MDA-MB-231 or 4T1 cells into the mammary fat pads of athymic or immune-competent BALB/c mice, respectively, followed by surgical excision of the primary tumour. Suppression of CIN had no effect on the efficiency of primary tumour implantation, and even enhanced primary tumour growth in the 4T1 model. However, in both models, suppression of CIN significantly reduced spontaneous metastasis and prolonged survival (Extended Data Fig. 3d, e).

We then assessed chromosome missegregation in the injected cells and in cells derived from primary tumours and metastatic colonies (Fig. 3a). We performed this analysis using MDA-MB-231 cells and two metastasis-competent xenografts (PDX) derived from patients with oestrogen receptor-positive (ER+) and triple-negative breast cancer (TNBC). Regardless of the CIN status of the injected cells, the majority of metastases were enriched for higher rates of chromosome missegregation, whereas cells derived from most primary tumours had significantly lower rates of CIN (Fig. 3b–d). For instance, when CIN-high cells (Fig. 3d, dnMCAK, blue bars) were injected into the mammary fat pad, chromosome missegregation rates decreased in the primary tumours (green bars) before increasing once more in metastases spontaneously arising within the same animal (orange bars).

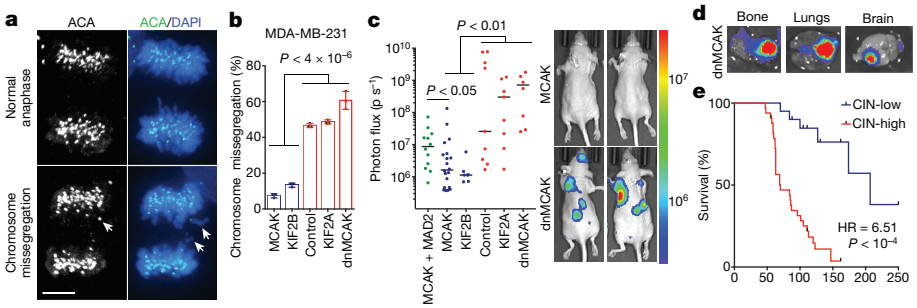


Figure 2 | CIN is a driver of metastasis. a, Anaphase cells stained for centromeres (ACA) and DNA (DAPI); scale bar, 5 μm. **b**, Chromosome missegregation in MDA-MB-231 cells expressing kinesin-13 proteins. Bars represent mean ± s.d., $n = 150$ cells. **c**, Whole animal bioluminescence (BLI) seven weeks after intracardiac injection of MDA-MB-231 cells. Left, bars represent the median and data points represent individual mice; $n = 12$ (MCAK + MAD2), 20 (MCAK), 7 (KIF2B), 9 (control), 9 (KIF2A), 8 (dnMCAK) mice. Right, representative images; colour scale shows photon flux. **d**, *Ex vivo* BLI of organs with metastases from MDA-MB-231 cells expressing dnMCAK; colour scale as in **c**. **e**, Disease-specific survival of mice injected with CIN-high ($n = 33$) or CIN-low ($n = 20$) MDA-MB-231 cells. Significance tested using two-sided *t*-test (**b**), two-sided Mann–Whitney test (**c**), and two-sided log-rank test (**e**); $n = 3$ (**a**, **b**) and 5 (**d**) independent experiments. Throughout the paper, pairwise comparisons between individual CIN-low and CIN-high conditions are smaller than the stated *P* value. HR, hazard ratio.

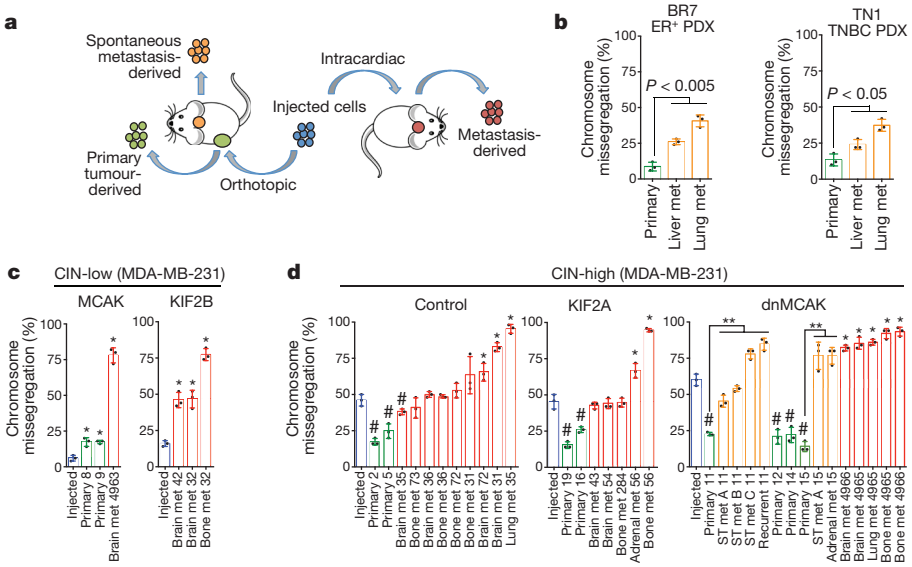


Figure 3 | Opposing roles for CIN in primary tumours and metastases. a, Experimental schema. **b–d**, Chromosome missegregation in injected cells (blue), cells derived from primary tumours (green), spontaneous metastases (met) (orange) or metastases arising after intracardiac injection (red). ST, soft tissue. Data shown as mean ± s.d., $n = 150$ cells, three independent experiments; $*P < 0.05$, $\#P < 0.05$, $**P < 0.05$ denote samples with higher or lower chromosome missegregation than the injected lines, and spontaneous metastases with higher missegregation than matched primary tumours, respectively. Significance tested using two-sided *t*-test (**b–d**).

CIN enriches for mesenchymal traits

Bulk RNA sequencing (RNA-seq) identified 1,584 genes that were differentially expressed between CIN-low and CIN-high MDA-MB-231 cells. Principal component analysis and unsupervised clustering accurately separated samples according to their CIN status. Metastasis-related and epithelial-to-mesenchymal transition (EMT) gene sets were relatively enriched in CIN-high cells. The top 23 differentially expressed genes in CIN-high cells (referred to as CIN signature) predicted distant metastasis-free survival (DMFS) in a meta-analysis[16] as well as a validation cohort[17] of patients with breast cancer, irrespective of tumour subtype, grade, or lymph node status (Extended Data Figs 4, 5).

RNA-seq of primary tumour-derived and metastasis-derived cells revealed pathways that were shared among metastases and CIN-high cells. However, metastases contained a large number of differentially upregulated EMT and inflammation-related genes that were disproportionately clustered on chromosome 1, signifying chromosome 1 -specific selection. Karyotype analysis revealed that the injected cell lines and most metastases had three copies of chromosome 1, whereas primary tumours consistently had two copies. Thus chromosome 1 loss is a recurrent event during primary tumour growth in this model (Extended Data Figs 4c–f, 5a–e).

We then performed single-cell RNA-seq (scRNA-seq) on three MDA-MB-231 cell lines—two CIN-low (KIF2B and MCAK) and one CIN-high (dnMCAK)—comprising a total of 6,821 cells. Clustering of single cells using EMT genes successfully classified most cells according to their CIN-status and revealed a fraction of cells (primarily CIN-high) that was highly enriched in mesenchymal markers (Fig. 4a). Unsupervised graph-based clustering, based on all expressed genes, identified 12 phenotypically distinct subpopulations. One subpopulation was defined by increased expression of genes involved in EMT and metastasis (referred to as subpopulation 'M') and was concomitantly enriched for CIN signature genes. This subpopulation comprised 45% of dnMCAK expressing cells compared to 6% of CIN-low cells (Fig. 4b and Extended Data Fig. 6a, b).

In agreement with the scRNA-seq data, CIN-high cells exhibited increased migratory and invasive behaviour *in vitro*, and displayed evidence of actin cytoskeletal reorganization, diffuse vimentin staining, and increased cytoplasmic and nuclear localization of β-catenin (Extended Data Figs 6c, d, 7a–d). As expected, MAD2 overexpression rescued invasion and migration of MCAK-expressing cells. Furthermore, the ability of KIF2B or MCAK overexpression to suppress invasion *in vitro* was dependent on the cell cycle, as the addition of thymidine after transient transfection of either protein abrogated this phenotype (Extended Data Figs 6f, 7e, f and Supplementary Fig. 2).

CIN generates cytosolic DNA

To better define CIN-responsive pathways, we performed a gene–gene Pearson correlation analysis using scRNA-seq data and identified two large gene modules: module 1 was characterized by proliferative and metabolic pathways, whereas module 2 comprised EMT and inflammation gene sets (Fig. 5a). There was a strong positive correlation between inflammation-related, CIN signature, and EMT genes in the scRNA-seq and bulk RNA-seq data (Figs 4b, 5b and Extended Data Fig. 4b, c).

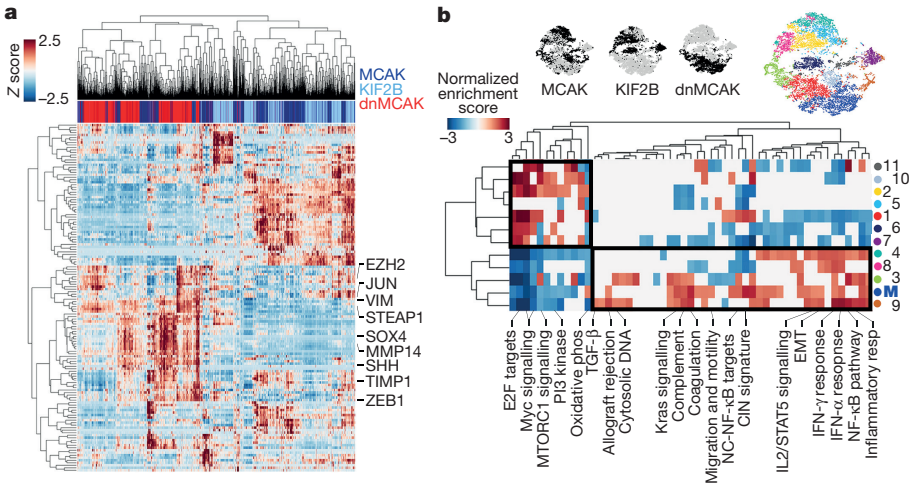


Figure 4 | CIN enriches for mesenchymal cell traits. a, Heat map showing expression of EMT genes in 6,821 MDA-MB-231 cells expressing MCAK, KIF2B, and dnMCAK. **b**, Top, t-stochastic neighbour embedding (tSNE) projection of all cells in **a**. Bottom, heat map of normalized enrichment score (NES) for pathways with an FDR $q < 0.05$ based on gene set enrichment analysis using one-sided weighted Smirnov–Kolmogorov test on subpopulations identified using Phenograph. Boxes outline populations with mutually exclusive transcriptional profiles.

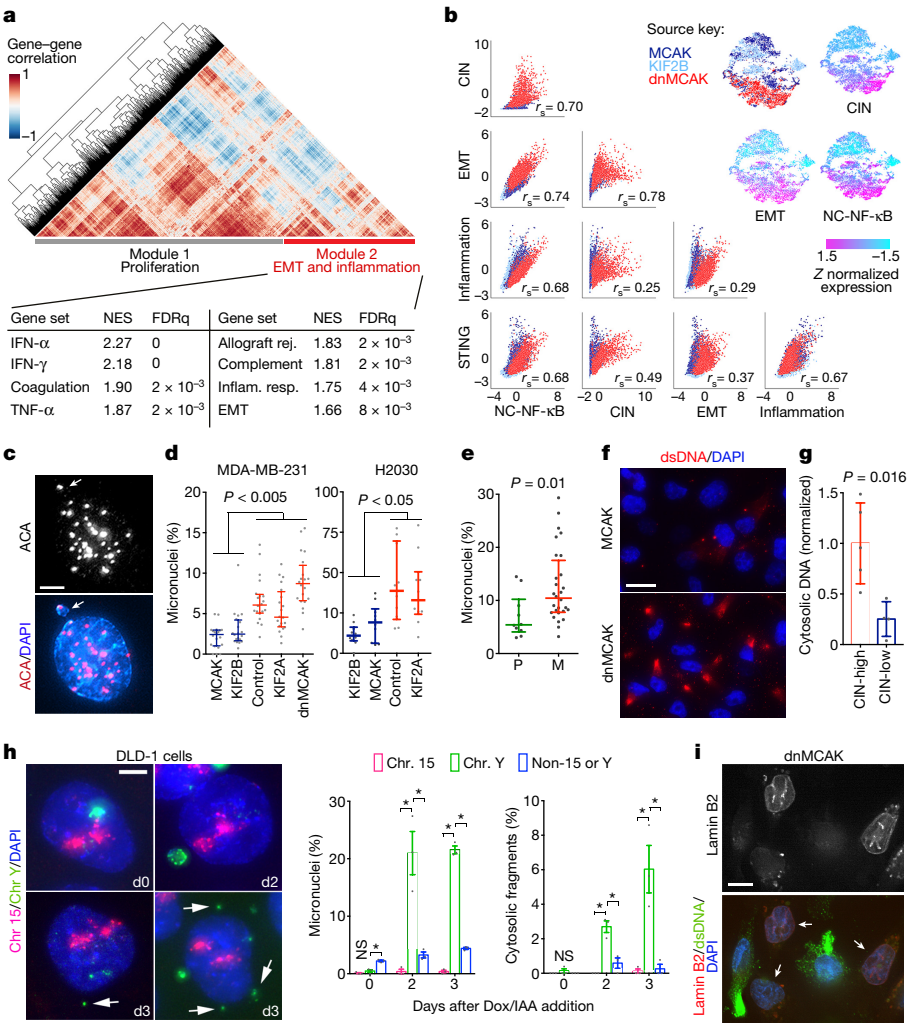


Figure 5 | CIN generates cytosolic DNA. **a**, Heat map showing gene–gene correlations in 6,821 cells and HALLMARK gene sets significantly enriched in module 2, one-sided weighted Smirnov–Kolmogorov test. **b**, Top right, normalized expression of key gene signatures in 6,821 MDA-MB-231 cells expressing MCAK, KIF2B, and dnMCAK. Bottom left, correlation plots for key gene signatures. **c**, A primary nucleus and a micronucleus stained for centromeres and DNA; scale bar 5 μm. **d**, **e**, Percentage of micronuclei cells expressing various kinesin-13 proteins (**d**) or in cells derived from 10 primary tumours and 28 metastases (**e**). $n = 20$ high-power fields per sample (**d**) or average values derived from 10 high-power fields per sample (**e**). **f**, MDA-MB-231 cells expressing MCAK and dnMCAK stained using DAPI and anti-dsDNA antibody; scale bar, 20 μm. **g**, Cytosolic-to-nuclear DNA ratios in CIN-high ($n = 5$) and CIN-low ($n = 4$) MDA-MB-231 ($n = 5$) and H2030 ($n = 4$) cells. **h**, Left, DLD-1 cells stained for DNA and hybridized to chromosome-specific FISH probes; scale bar, 5 μm. Right, percentage of cells containing micronuclei or small cytosolic DNA fragments ($n > 500$ cells per condition). **i**, MDA-MB-231 cells stained using DAPI, anti-dsDNA antibody, or mCherry–lamin B2 (arrow); scale bar, 10 μm. Data shown as median ± interquartile range (**d**, **e**), mean ± s.d. (**g**), mean ± s.e.m. (**h**). $n = 6$ (**c**, **d**), 3 (**e**, **f**, **h**), 1 (**g**) and 2 (**i**) independent experiments. Significance tested using two-sided Mann–Whitney test (**d**, **e**) or two-sided *t*-test (**g**, **h**); $^*P < 0.05$.

The induction of inflammatory pathways in response to CIN was unexpected and was reminiscent of a viral infection. We investigated whether CIN might introduce genomic DNA into the cytosol, thereby eliciting cellular responses normally reserved for anti-viral immunity[18,19]. The exposure of genomic DNA to the cytosol can result from either primary nuclear or micronuclear envelope ruptures[20–24]. We performed live-cell imaging using a GFP reporter with a nuclear localization signal (NLS–GFP)[25] and found no correlation between CIN and the frequency of NLS–GFP leakage into the cytosol in unconfined conditions. There was even a trend towards more efficient primary nucleus repair in CIN-high cells. CIN-high nuclei ruptured more frequently only during confined migration, and this was primarily attributed to their increased ability to go through a larger number of small constrictions (Extended Data Fig. 7g–j) that mimic confined migration during metastasis[25].

Instead, CIN-high cells and those derived from metastases exhibited a higher preponderance of micronuclei than did CIN-low or primary tumour-derived cells, respectively (Fig. 5c–e and Extended Data Fig. 8a–c). To test whether the presence of rupture-prone[20] micronuclei correlated with increased cytosolic DNA, we stained cells using two different anti-dsDNA antibodies after selective plasma membrane permeabilization and found increased cytosolic dsDNA and single-stranded DNA (ssDNA) in CIN-high cells. The dsDNA signal, which was distinct from mitochondrial staining, disappeared after treatment with double-strand-specific—but not single-strand-specific—nuclease and after overexpression of DNASE2, confirming the specificity of these antibodies (Fig. 5f and Extended Data Fig. 8d–h). Quantification of dsDNA levels after subcellular fractionation revealed a fourfold reduction in cytosolic DNA in CIN-low cells compared to CIN-high cells (Fig. 5g). Whole-genome sequencing of subcellular fractions at 30× coverage confirmed the genomic origin of cytosolic DNA (data not shown).

To determine whether missegregated chromosomes provide a source of cytosolic DNA, we used an inducible Y-chromosome-specific missegregation system established in chromosomally stable DLD-1 colorectal cancer cells[26]. Whole-chromosome fluorescence *in situ* hybridization (FISH) probes targeting the Y chromosome or an independent autosome (chromosome 15) revealed selective incorporation of the Y chromosome into micronuclei two days after chromosome missegregation induced by doxycycline and auxin (Dox/IAA) treatment. Notably, Y-chromosome-specific fragments were found dispersed within the cytosol 2–3 days after Dox/IAA addition, whereas the control autosome remained confined to the nucleus (Fig. 5h), demonstrating that cytosolic DNA is generated from chromosomes undergoing high rates of missegregation

Suppression of micronuclear envelope rupture by mCherry–lamin B2 overexpression[20] reduced cytosolic dsDNA staining without influencing chromosome segregation errors. Accordingly, such overexpression reduced metastasis after intracardiac or tail vein injection of MDA-MB-231 cells (Fig. 5i and Extended Data Fig. 3g, h).

Metastasis from cytosolic DNA response

In chromosomally stable cells, cytosolic dsDNA is scarce and is sensed by the cGAS–STING pathway[19], leading to induction of type I interferon stimulated genes (ISGs)[22,23,27]. Indeed, induced missegregation of the Y chromosome led to the upregulation of OAS2, an ISG, and increased interferon-β production by DLD-1 cells

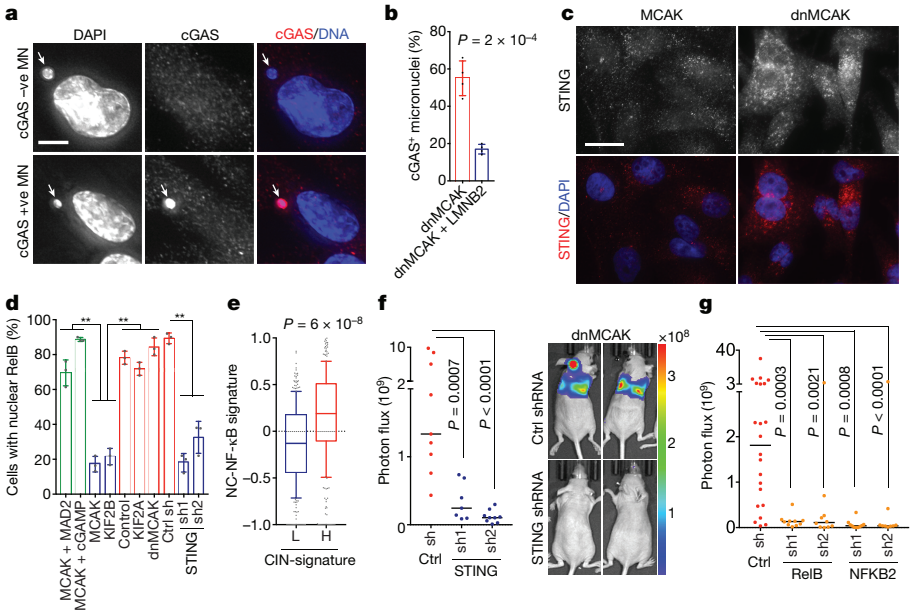


Figure 6 | Metastasis from a cytosolic DNA response. a, MDA-MB-231 cells stained for DNA and cGAS; scale bar, 5 μm. **b**, Percentage of micronuclei with cGAS localization (cGAS$^+$), $n = 200$ cells. **c**, Cells stained for DNA and STING; scale bar, 20 μm. **d**, Percentage of MDA-MB-231 cells with nuclear RELB, $n = 150$ cells. **e**, Average *z*-normalized expression of CIN-responsive noncanonical NF-κB target genes in breast cancer patients with low (<30th percentile, $n = 330$) or high (>30th percentile, $n = 332$) CIN gene expression signature. **f**, **g**, Photon flux (p s^{-1}) of whole animals after intracardiac injection with MDA-MB-231 cells expressing control shRNA, STING shRNA (**f**), RELB shRNA (**g**) or NFKB2 shRNA (**g**); $n = 9$, 7, and 9 mice for the control, STING shRNA1, and STING shRNA2 groups, respectively (**f**); $n = 22$, 10, 10, 10, and 9 mice for the control, RELB shRNA1, RELB shRNA2, NFKB2 shRNA1, and NFKB2 shRNA2 groups, respectively (**g**). Data shown as mean ± s.d. (**b**, **d**), median ± interquartile range with bars spanning 10th–90th percentiles (**e**), median (**f**, **g**); significance tested using two-sided *t*-test (**b**, **d**), two-sided Mann–Whitney test (**e**–**g**); $n = 4$ (**a**, **b**) and 3 (**c**, **d**) independent experiments.

(Extended Data Fig. 9f, g). It is unclear how chromosomally unstable cancer cells cope with the constant presence of cytosolic DNA. We found notable localization of cGAS to approximately half of all micronuclei, as previously observed[22,23,27]. Impeding micronuclear rupture through lamin B2 overexpression[20] significantly diminished the relative fraction of cGAS$^+$ micronuclei (Fig. 6a, b). Furthermore, CIN-high cells exhibited increased levels and perinuclear localization of STING, congruent with pathway activation (Fig. 6c).

Notably, there was no evidence for robust activation of downstream canonical NF-κB or type I interferon signalling in CIN-high cells, as evidenced by the lack of a significant increase in p65 or IRF3 phosphorylation, absence of p65 or IRF3 nuclear translocation, undetectable levels of interferon-β, and failure to induce ISGs (Extended Data Figs 8i, j, 9), in line with previous observations[28–30].

Cytosolic DNA, however, can activate the noncanonical NF-κB pathway in a STING-dependent and TBK1-independent manner[18]. We found evidence for noncanonical NF-κB activation in CIN-high cells, as revealed by lower levels of the precursor protein p100, a trend towards higher ratios of p52 and phosphorylated p100 relative to total p100, and reduced levels of the noncanonical NF-κB pathway inhibitor TRAF2[31] (Extended Data Fig. 8i, j). Given the subtle differences seen at the protein level, we assessed the nuclear localization of RELB, the binding partner of p52, and observed increased nuclear localization in CIN-high cells. This was often accompanied by cytosolic staining, indicative of chronic pathway activation. STING depletion reduced nuclear localization of RELB and led to downregulation of EMT and inflammatory pathways, whereas the addition of cGAMP or overexpression of MAD2 increased nuclear RELB in MCAK-expressing cells (Fig. 6d, Extended Data Figs 4e, 9c–e).

Bulk RNA-seq data identified a number of noncanonical NF-κB target genes that were upregulated in response to CIN (CIN-responsive NC-NF-κB genes). There was a robust correlation between the CIN signature, STING, and the CIN-responsive NC-NF-κB genes in scRNA-seq data, in contrast to a weaker correlation between CIN and type I interferon targets (Fig. 5b and Extended Data Fig. 6e). Similarly, RNA-seq data from primary breast cancer in the TCGA database demonstrated increased expression of CIN-responsive NC-NF-κB genes in tumours with higher levels of CIN signature gene products (Fig. 6e), and higher expression of key regulators of the noncanonical NF-κB pathway or its CIN-responsive target genes was associated with shorter DMFS and disease-free survival in breast and lung cancers. Conversely, increased expression of canonical NF-κB or type I interferon regulatory factors was associated with an improved prognosis (Extended Data Fig. 10a).

cGAS activation by cancer cells has been invoked in brain metastasis through a tumour cell-non-autonomous mechanism[32]. We found that STING and downstream noncanonical NF-κB activity mediate metastasis in a tumour cell-autonomous fashion, as evidenced by reduction in metastatic dissemination, lifespan extension, and reduction in *in vitro* and *in vivo* invasion of CIN-high cells depleted of STING, RELB, or p100 (encoded by *NFKB2*). Conversely, the addition of cGAMP increased invasion and migration of CIN-low cells (Fig. 6f, g and Extended Data Figs 6f, 9j, 10b–d). These findings are in line with reported roles for the noncanonical NF-κB pathway in EMT, cellular invasion, and metastasis[33–35]. The benefits of the noncanonical pathway may justify the scarcity of inactivating mutations in cGAS and STING among breast and lung cancers (Extended Data Fig. 10e).

Discussion

Our work reveals an unexpected link between CIN, chronic activation of cytosolic DNA sensing pathways, and metastasis. In addition to fuelling karyotypic heterogeneity that serves as a substrate for natural selection, ongoing chromosome missegregation is required to replenish cytosolic DNA pools and to maintain cells in a pro-metastatic state. Consequently, suppression of CIN reduces metastasis even in highly aneuploid cells. The repercussions of STING activation are context-dependent and range from senescence to tumorigenesis[21–23,27,28,30].

Given that chromosomally unstable cells are awash with cytosolic DNA, our results raise the possibility that by suppressing downstream type I interferon signalling[30] and instead upregulating the alternative NF-κB pathway, such cells have substituted a lethal epithelial response to inflammation with that of myeloid-derived cells[36,37], thereby engaging in some form of immune mimicry. Restoration of normal responses to inflammation would constitute a viable therapeutic strategy to target chromosomally unstable cells.

The emergence, and subsequent tolerance, of CIN represents an important bottleneck during tumour evolution[38–40]. We found that CIN induces a transcriptional shift from a proliferative and highly metabolic state, ideally suited for primary tumour growth, to a mesenchymal state associated with upregulation of inflammatory pathways (Figs 4b, 5a). These two mutually exclusive states, which were recently observed in a pan-cancer genomic analysis of metastatic tumours[41], are likely to account for the reversibility in chromosome missegregation rates seen in primary tumours and metastases, and provide an explanation for the negative effect of aneuploidy during early tumorigenesis[38,40,42]. These findings also lead us to suggest that CIN drives the subset of human metastases characterized by EMT and inflammation[41].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 April; accepted 6 December 2017.
Published online 17 January 2018.

1. Jamal-Hanjani, M. *et al.* Tracking the evolution of non-small-cell lung cancer. *N. Engl. J. Med.* **376,** 2109–2121 (2017).
2. Turajlic, S. & Swanton, C. Metastasis as an evolutionary process. *Science* **352,** 169–175 (2016).
3. Thompson, S. L. & Compton, D. A. Examining the link between chromosomal instability and aneuploidy in human cells. *J. Cell Biol.* **180,** 665–672 (2008).
4. Cimini, D. *et al.* Merotelic kinetochore orientation is a major mechanism of aneuploidy in mitotic mammalian tissue cells. *J. Cell Biol.* **153,** 517–527 (2001).
5. Bakhoum, S. F. *et al.* The mitotic origin of chromosomal instability. *Curr. Biol.* **24,** R148–R149 (2014).
6. Bakhoum, S. F. *et al.* Numerical chromosomal instability mediates susceptibility to radiation treatment. *Nat. Commun.* **6,** 5990 (2015).
7. Bakhoum, S. F., Thompson, S. L., Manning, A. L. & Compton, D. A. Genome stability is ensured by temporal control of kinetochore–microtubule dynamics. *Nat. Cell Biol.* **11,** 27–35 (2009).
8. Burrell, R. A. *et al.* Replication stress links structural and numerical cancer chromosomal instability. *Nature* **494,** 492–496 (2013).
9. Brastianos, P. K. *et al.* Genomic characterization of brain metastases reveals branched evolution and potential therapeutic targets. *Cancer Discov.* **5,** 1164–1177 (2015).
10. Mitelman, F., Johansson, B. & Mertens, F. *Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer* https://cgap.nci.nih.gov/Chromosomes/Mitelman (2017).
11. Chung, C. H. *et al.* Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. *Cancer Cell* **5,** 489–500 (2004).
12. Moore, A. T. *et al.* MCAK associates with the tips of polymerizing microtubules. *J. Cell Biol.* **169,** 391–397 (2005).
13. Ems-McClung, S. C. & Walczak, C. E. Kinesin-13s in mitosis: key players in the spatial and temporal organization of spindle microtubules. *Semin. Cell Dev. Biol.* **21,** 276–282 (2010).
14. Waterman-Storer, C. M., Worthylake, R. A., Liu, B. P., Burridge, K. & Salmon, E. D. Microtubule growth activates Rac1 to promote lamellipodial protrusion in fibroblasts. *Nat. Cell Biol.* **1,** 45–50 (1999).
15. Sotillo, R. *et al.* Mad2 overexpression promotes aneuploidy and tumorigenesis in mice. *Cancer Cell* **11,** 9–23 (2007).
16. Győrffy, B. *et al.* An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res. Treat.* **123,** 725–731 (2010).
17. Hatzis, C. *et al.* A genomic predictor of response and survival following taxane–anthracycline chemotherapy for invasive breast cancer. *J. Am. Med. Assoc.* **305,** 1873–1881 (2011).
18. Abe, T. & Barber, G. N. Cytosolic-DNA-mediated, STING-dependent proinflammatory gene induction necessitates canonical NF-κB activation through TBK1. *J. Virol.* **88,** 5328–5341 (2014).
19. Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. Cyclic GMP–AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339,** 786–791 (2013).
20. Hatch, E. M., Fischer, A. H., Deerinck, T. J. & Hetzer, M. W. Catastrophic nuclear envelope collapse in cancer cell micronuclei. *Cell* **154,** 47–60 (2013).
21. Ahn, J. *et al.* Inflammation-driven carcinogenesisis mediated through STING. *Nat. Commun.* **5,** 5166 (2014).
22. Harding, S. M. *et al.* Mitotic progression following DNA damage enables pattern recognition within micronuclei. *Nature* **548,** 466–470 (2017).
23. Mackenzie, K. J. *et al.* cGAS surveillance of micronuclei links genome instability to innate immunity. *Nature* **548,** 461–465 (2017).
24. Lan, Y. Y., Londoño, D., Bouley, R., Rooney, M. S. & Hacohen, N. Dnase2a deficiency uncovers lysosomal clearance of damaged nuclear DNA via autophagy. *Cell Reports* **9,** 180–192 (2014).
25. Denais, C. M. *et al.* Nuclear envelope rupture and repair during cancer cell migration. *Science* **352,** 353–358 (2016).
26. Ly, P. *et al.* Selective Y centromere inactivation triggers chromosome shattering in micronuclei and repair by non-homologous end joining. *Nat. Cell Biol.* **19,** 68–75 (2017).
27. Yang, H., Wang, H., Ren, J., Chen, Q. & Chen, Z. J. cGAS is essential for cellular senescence. *Proc. Natl Acad. Sci. USA* **114,** E4612–E4620 (2017).
28. Stetson, D. B., Ko, J. S., Heidmann, T. & Medzhitov, R. Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell* **134,** 587–598 (2008).
29. Lau, L., Gray, E. E., Brunette, R. L. & Stetson, D. B. DNA tumor virus oncogenes antagonize the cGAS–STING DNA-sensing pathway. *Science* **350,** 568–571 (2015).
30. Dou, Z. *et al.* Cytoplasmic chromatin triggers inflammation in senescence and cancer. *Nature* **550,** 402–406 (2017).
31. Sun, S.-C. Controlling the fate of NIK: a central stage in noncanonical NF-κB signaling. *Sci. Signal.* **3,** pe18 (2010).
32. Chen, Q. *et al.* Carcinoma–astrocyte gap junctions promote brain metastasis by cGAMP transfer. *Nature* **533,** 493–498 (2016).
33. Wang, X. *et al.* Oestrogen signalling inhibits invasive phenotype by repressing RelB and its target BCL2. *Nat. Cell Biol.* **9,** 470–478 (2007).
34. Wang, J., Yi, S., Zhou, J., Zhang, Y. & Guo, F. The NF-κB subunit RelB regulates the migration and invasion abilities and the radio-sensitivity of prostate cancer cells. *Int. J. Oncol.* **49,** 381–392 (2016).
35. Yang, C. *et al.* Antagonism of inhibitor of apoptosis proteins increases bone metastasis via unexpected osteoclast activation. *Cancer Discov.* **3,** 212–223 (2013).
36. Gulen, M. F. *et al.* Signalling strength determines proapoptotic functions of STING. *Nat. Commun.* **8,** 427 (2017).
37. Kis-Toth, K., Szanto, A., Thai, T. H. & Tsokos, G. C. Cytosolic DNA-activated human dendritic cells are potent activators of the adaptive immune response. *J. Immunol.* **187,** 1222–1234 (2011).
38. Sheltzer, J. M. *et al.* Single-chromosome gains commonly function as tumor suppressors. *Cancer Cell* **31,** 240–255 (2017).
39. Weaver, B. A. A., Silk, A. D., Montagna, C., Verdier-Pinard, P. & Cleveland, D. W. Aneuploidy acts both oncogenically and as a tumor suppressor. *Cancer Cell* **11,** 25–36 (2007).
40. Rowald, K. *et al.* Negative selection and chromosome instability induced by Mad2 overexpression delay breast cancer but facilitate oncogene-independent outgrowth. *Cell Reports* **15,** 2679–2691 (2016).
41. Robinson, D. R. *et al.* Integrative clinical genomics of metastatic cancer. *Nature* **548,** 297–303 (2017).
42. Sheltzer, J. M. A transcriptional and metabolic signature of primary aneuploidy is present in chromosomally unstable cancer cells and informs clinical prognosis. *Cancer Res.* **73,** 6401–6412 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Massagué, R. Benezra, X. Cai, J. Leeman, M. Bakhoum, B. Hopkins and D. Landau for feedback. Grant support: S.F.B. (DoD Breast Cancer Research Breakthrough Award (BCRBA) W81XWH-16-1-0315, Elsa Pardee Foundation, MSKCC Cytogenetics Core (P30-CA008748) and Core (P30-CA008748) grants); B.N. (NSF Graduate Research Fellowship DGE1257284); J.L. (NIH R01-HL082792, U54-CA210184, DoD BCRBA BC150580, NSF CBET-1254846); P.L. (NCI K99-CA218871); G.G. (DoD BCRBA W81XWH-16-1-0316); L.C.C. (NIH R35-CA197588, U54-CA210184, Breast Cancer Research Foundation, Gray Foundation Basser Initiative).

Author Contributions S.F.B. and L.C.C. conceived the project; S.F.B., B.N., J.-A.C. and R.K.S. performed animal experiments; V.K.R. derived PDX models; P.L. performed experiments using DLD-1 cells; S.F.B., B.N., J.A.C., R.K.S., M.D., S.V. and Q.L. performed immunostaining, immunoblotting, microscopy and qPCR; P.S. performed live-cell imaging; T.B.K.W., N.J.B. and N.M. analysed matched primary tumour–metastasis data; N.K.T., C.H.C. and S.F.B. analysed the HNSCC data; S.F.B. analysed breast cancer karyotype data; C.P. performed histological analysis; C.S., K.C. and G.N. performed cytogenetic analysis; A.M.L. analysed scRNSseq data; M.L. and S.F.B. analysed survival data; C.J.M. analysed bulk RNAseq data; M.I. analysed whole-genome sequence data; G.G., M.L., Q.L., J.H.H., O.E., C.H.C., N.Y.L., D.P., D.W.C., S.N.P., J.L., C.S. and L.C.C. assisted with data interpretation. All authors contributed to the writing and editing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to L.C.C. (lcantley@med.cornell.edu).

Reviewer Information *Nature* thanks N. Gekara, J. van Deursen and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Genomic analysis of primary tumour–metastasis matched pairs. Whole-exome DNA sequence data from 79 brain metastases with matched primary tumour and normal tissue[9] were downloaded from the database of Genotypes and Phenotypes (dbGAP) and processed as described[43] to derive allele-specific segmented DNA copy number data for each sample. The weighted genome instability index (wGII), which describes the proportion of the genome that was classified as aberrant relative to tumour ploidy, was determined as described[8].

Mitelman database analysis. All available breast adenocarcinoma cases in the Mitelman database[10] were analysed. The primary literature was reviewed to determine the source of the sample (primary tumour or metastasis). When clonal karyotype was reported as a range, the average value was used for that given clone. Karyotype aberrations included structural aberrations as well as numerical deviations from the overall karyotype of the clone.

Analysis of chromosome segregation in head and neck squamous cell carcinoma. We analysed primary tumour specimens from 60 patients with head and neck squamous cell carcinoma (HNSCC)[11]. Haematoxylin and eosin-stained primary tumour samples of sufficient quality for high-resolution microscopy analysis were available for forty patients. Analysis was restricted to cells fixed while undergoing anaphase, as previously described[44,45]. Chromosome missegregation was defined by the presence of haematoxylin staining between the remaining segregating chromosomes during anaphase and was reported as the percentage of cells undergoing anaphase with evidence of chromosome missegregation, as previously described[44]. Clinical lymph node status was defined on the basis of clinical examination or radiographic evidence of lymph node tumour involvement[11].

Single-cell karyotyping. Cultures were treated with colcemid at a final concentration of 0.1 μg ml^{-1}. Following 45 min incubation at 37 °C, the cultures were trypsinized, resuspended in pre-warmed 0.075 M KCl, incubated for an additional 10 min at 37 °C and fixed in methanol:acetic acid (3:1). The fixed cell suspension was then dropped onto slides, stained in 0.08 μg ml^{-1} DAPI in 2×SSC for 5 min and mounted in antifade solution (Vectashield, Vector Labs). Metaphase spreads were captured using a Nikon Eclipse E800 epifluorescence microscope equipped with GenASI Cytogenetic suite (Applied Spectral Imaging). For each sample a minimum of 20 inverted DAPI-stained metaphases were fully karyotyped and analysed according to the International System of Human Cytogenetic Nomenclature (ISCN) 2013.

FISH analysis. FISH analysis was performed on fixed cells prepared for single-cell karyotyping. Based on karyotype data, four chromosomes were selected to further evaluate numerical instability. Probes specific for centromere 3 (red), centromere 4 (orange) and centromere 9 (green) were purchased form Abbott-Vysis. The chromosome 6 centromeric probe was home-brew (PAC clone P308; labelled with green dUTP, MSKCC Molecular Cytogenetics Core Facility). Probe labelling, hybridization, post-hybridization washing, and fluorescence detection were performed according to standard laboratory procedures[26]. For each probe, three normal peripheral blood samples (GM07535, GM06875 and GM00558), obtained from Coriell Institute, were also analysed to derive cut-off values (false positive).

RHOA and RAC1 pull-down assays. The activity of RHOA and RAC1 was determined using bead-based pull-down assay kits (Cytoskeleton, RHOA: BK036S, RAC1: BK035S). Cells were lysed on the tissue culture dish and rapidly snap frozen for storage until further processing. cGAMP was added for 18 h before lysis. In addition to His-tagged RHOA and RAC1, the positive and negative controls were total cell lysates supplemented with non-hydrolysable GTP or GDP, respectively.

Cell culture. Cell lines were purchased from the American Type Culture Collection (ATCC). Tumour (MDA-MB-231, 4T1, HEK293, and H2030), cells were cultured in DMEM or RPMI (4T1) supplemented with 10% FBS and 2 mM L-glutamine in the presence of penicillin (50 U ml^{-1}) and streptomycin (50 μg ml^{-1}). All cells tested negative for mycoplasma. Cell confluence was measured using IncuCyte live-cell analysis system (Essen Bioscience).

Immunofluorescence microscopy. Cell fixation and antibody staining were performed as previously described[7]. In brief, cells were fixed with ice-cold (−30 °C) methanol for 15 min (when staining for centromeres, centrosomes, cGAS, Vimentin, β-actin, IRF3, or α-tubulin) or 4% paraformaldehyde (when staining for RELB, p65, STING, ssDNA, dsDNA, COXIV, or β-catenin). Subsequently, cells were permeabilized using 1% triton for 4 min. See Supplementary Table 1 for antibody information. For selective plasma membrane permeabilization used for cytosolic dsDNA and ssDNA staining, cells were treated with 0.02% saponin for 5 min after fixation. For single-stranded (Thermo Fisher Scientific, FEREN0321) and double stranded (Life Technologies, EN0771)-specific nuclease treatment, cells were also permeabilized with 0.02% saponin for 2 min and treated with either nuclease for 10 min before fixation using 4% paraformaldehyde. TBS–BSA was used as a blocking agent during antibody staining. DAPI was added together with secondary antibodies. Cells were mounted with Prolong Diamond Antifade Mountant (Life Technologies, P36961). cGAMP (Invivogen, tlrl-nacga23) was transfected into cells at a concentration of 4 μg ml^{-1} using lipofectamine 2000 that was added for 3–4 h and then replaced with regular serum-containing medium.

Immunoblotting. Cells were pelleted and lysed using RIPA buffer. Protein concentration was determined using BCA protein assay and 20–30 μg total protein was loaded in each lane. Proteins were separated by gradient SDS–PAGE and transferred to PVDF or nitrocellulose membranes. Full blots are shown in Supplementary Fig. 1. See Supplementary Table 2 for antibody information. Band intensities on immunoblots were obtained using ImageJ (https://imagej.nih.gov/ij/) or the LI-COR Odyssey software, normalized to loading control, and background was subtracted. Ratios were normalized to control cells. Interferon-β levels from conditioned medium were measured using the Human IFN beta Array 1-plex (Eve Technologies, HIFNB-01-31).

Y chromosome missegregation and FISH. Flp-In T-REx DLD-1 cells were engineered to express the TIR1 auxin-dependent plant E3 ligase, an auxin-inducible degron (AID)-tagged CENP-A modified at the endogenous allele (CENP-A$^{AID/-}$), and a doxycycline-inducible CENP-A^{C-H3} rescue gene integrated into the Flp-In locus as previously described[26]. Cells (4.0×10^4) were seeded into 4-well chamber slides and treated with doxycycline (DOX, Sigma-Aldrich) and indole-3-acetic acid (IAA, Sigma-Aldrich) for up to 3 days to induce Y chromosome missegregation and micronuclei. Slides were washed in PBS, fixed in 3:1 methanol:acetic acid for 15 min at room temperature, and dehydrated with 80% ethanol. Chromosome paint FISH probes targeting chromosomes Y and 15 (MetaSystems) were mixed at equal ratios, applied to cells, sealed with a coverslip, and co-denatured at 75 °C for 2 min followed by overnight hybridization at 37 °C in a humidified chamber. Slides were washed in 0.4× saline–sodium citrate (SSC) buffer for 2 min at 72 °C, followed by a 30 s wash in 2× SCC, 0.05% Tween-20 buffer at room temperature. Cells were counterstained with DAPI and captured on a DeltaVision Elite (GE Healthcare) microscope system at 60× magnification (25 × 0.2 μm *z*-stacks) followed by image deconvolution and maximum intensity quick projection.

Knockdown and overexpression constructs. Luciferase expression was achieved using pLVX plasmid (expressing tdTomato) and cells stably expressing luciferase were selected using hygromycin and sorted for tdTomato expression. Kinesin-13 family protein expression was achieved using plasmid (pEGFP) transfection or lentiviral (pLenti-GIII-CMV-GFP-2A-Puro) expression where cells were selected using G418 (0.5 mg ml^{-1}) or puromycin (5 μg ml^{-1}), respectively. DNASE2 overexpression was achieved using a pLenti-GIII-CMV-RFP-2A-Puro plasmid with puromycin used for selection. Plasmids containing kinesin-13 proteins or lamin B2 (pQCXIB–mCherry–lmnb2) constructs were offered by the Compton and Hetzer Laboratories, respectively. Blasticidin was used to select for *lmnb2*-expressing cells at 10 μg ml^{-1}. All other plasmids were purchased from Applied Biological Materials (https://www.abmgood.com/). Stable knockdown of STING, NFKB2, RELB, and cGAS were achieved using shRNAs in pRRL (SGEP or SGEN) plasmids obtained from the MSKCC RNA Interference Core. Two to four distinct shRNA hairpins were screened per target. Targeted shRNA sequences are listed in Supplementary Table 3. To visualize primary nuclear rupture, cells were stably modified with a retroviral construct expressing both NLS–GFP[25] and H2B–Tdtomato (3×NLScopGFP-P2A-H2BtdTomato-IRES-puromycin). Cells were cultured for 24 h after viral transduction before selection with 1 μg ml^{-1} puromycin and subsequently sorted to select for NLS–GFP and H2B–TdTomato expression.

Cell migration in microfluidic devices. Microfluidic migration devices with precisely defined constrictions were prepared as described previously[25,46]. Devices were coated with 50 μg ml^{-1} type-I rat tail collagen (BD Biosciences) in 0.02 M acetic acid overnight at 4 °C. Approximately 80,000 cells were seeded (in DMEM supplemented with 10% FBS and 1% PenStrep) per migration chamber. Devices were placed in a tissue-culture incubator (37 °C) for 5–6 h to allow the cells to adhere. Subsequently, the medium was changed to phenol-red free Leibovitz L15 medium supplemented with 10% FBS and 1% PenStrep before the device was mounted on an inverted microscope (Zeiss Observer Z1) equipped with a temperature-controlled stage (37 °C) for live-cell imaging. The medium reservoirs of the device were covered with glass coverslips to minimize evaporation during live-cell imaging. Cells were imaged for 14–16 h at 10-min intervals with a CCD camera (Photometrics CoolSNAP KINO) using a Zeiss 20×/NA 0.8 air objective. Acquired image sequences were analysed for nuclear rupture frequency, duration, and transit time of cells through 1 × 5-μm^2, 2 × 5-μm^2, and 15 × 5-μm^2 constrictions using Zen software (Zeiss) and a custom-written MATLAB 2016a script for automated image analysis.

Animal studies. Animal experiments were performed in accordance with protocols approved by the Weill Cornell Medicine Institutional Animal Care and Use Committee. For disease-specific survival in MDA-MB-231 experiments, power analysis indicated that ten mice per group would be sufficient to detect a difference

at relative hazard ratios of <0.2 or >5 with 80% power and 95% confidence, given a median disease-specific survival of 3 months in the control group and a total follow up period of 250 days. For the 4T1 experiments, power analysis indicated that ten mice per group would be sufficient to detect a difference at relative hazard ratios of <0.25 or >4.0 with 80% power and 95% confidence, given a median survival of 58 days in the control group and a total follow up period of 180 days. There was no need to randomize animals. Investigators were not blinded to group allocation. Intracardiac injection was performed as previously described[32]. In brief, cells were trypsinized and washed with PBS and 1×10^5 cells (in 100 µl PBS) were injected into the left cardiac ventricle of female athymic 6–7-week-old athymic nude (*nu*/*nu*) mice (Jackson Laboratory strain 002019). Cells (2×10^5) were injected into the tail vein cohort of animals. Mice were then immediately injected with D-luciferin (150 mg kg^{-1}) and subjected to bioluminescence imaging (BLI) using tan IVIS Spectrum Xenogen instrument (Caliper Life Sciences) to ensure systemic dissemination of tumour cells. Metastatic burden was measured 5–7 weeks after injection using BLI. BLI images were analysed using Living Image Software v.2.50. Disease-specific survival endpoint was met when the mice died or met the criteria for euthanasia under the IACUC protocol and had radiographic evidence of metastatic disease. For orthotopic tumour implantation, 2.5×10^5 cells in 50 µl PBS were mixed 1:1 with Matrigel (BD Biosciences) and injected into the fourth mammary fat pad. Only one tumour was implanted per animal. MDA-MB-231 primary tumours were surgically excised before they reached ~1.5 cm in the largest dimension (which was the maximum allowable under our IACUC protocol) and metastatic dissemination was assessed using BLI imaging at 1–3-week intervals for up to 30 weeks. The distant metastasis-free survival endpoint was met when BLI signal was seen outside the site of primary tumour transplantation. 4T1 tumours were excised 8–9 days after implantation. To derive short-term culture from primary tumours and metastases, anesthetized mice (isofluorane) were imaged then killed. *Ex vivo* BLI was subsequently performed on removed organs to define the precise location of the metastatic lesion. Primary tumours and metastases were subsequently mechanically dissociated and cultured in DMEM with selection medium (G418 or hygromycin) to select for tumour cells and exclude host cells. All subsequent assays (karyotyping, RNA-seq, immunofluorescence, and subcellular fractionation) were performed after a single passage from the primary sample. To assess chromosome missegregation from primary tumour-derived and metastasis-derived cells, we performed high-resolution immunofluorescence analysis on passage no. 1 cells, staining for DNA (DAPI) and centromeres (ACA). The presence of cells with DNA or centromere staining in the middle of the anaphase plate was taken as evidence of chromosome missegregation.

Patient-derived xenograft assays. Patient-derived xenograft (PDX) models of human metastatic breast cancers were generated by transplanting the freshly obtained surgically excised tumour specimens from patients who had given consent under the IRB approved protocol (MSKCC IRB #97-094) in female NOD/SCID/IL2Rγ^{null} (NSG) mice (Jackson Laboratories strain 005557). All relevant ethical regulations were followed. The oestrogen receptor-positive PDX was derived from breast cancer metastatic to the bone. The TNBC PDX was established from an axillary lymph node metastasis from a patient with inflammatory breast cancer. PDXs were maintained for a maximum of three serial passages. In brief, freshly obtained tumour tissue specimens were either directly transplanted in the mammary fat-pad of the mice or minced into 1–2-mm pieces in serum-free MEM medium with nonessential amino acids (Cat. No. 41500018, Thermo Fisher Scientific) transduced with lentiviral vectors expressing either GFP–luciferase or pUltra-Chili–Luc plasmid (Addgene plasmid: 48688) followed by transplantation into mice. Typically PDX tumour growth became evident during the first 1–3 weeks after engrafting and tumours continued to grow for an additional 4–8 weeks. Primary tumour growth and metastases were followed using BLI or spectrum CT imaging. At the time of removal of primary tumours and metastases, we derived primary cell cultures directly from primary tumours as well as lung and liver metastases. In brief, 500 mg of fresh bulk tumour tissue was chopped into 1–2-mm^3 pieces and incubated in Accutase (AT104; Innovative Cell Technologies) for cell detachment and separation over 1–2 h. The dissociated tissues were sieved through 100-µm cell strainers and the cells were pelleted by centrifugation at 1,200 r.p.m. The pellets were washed and resuspended in the above MEM buffer with 3% FBS. Cells were analysed for chromosome missegregation after one passage.

RNA sequencing and analysis. Bulk RNA was extracted from cells using the QIAShredder (Qiagen, 79654) and the RNA extraction kit (Qiagen, 74106) and sequenced using HiSeq2500 or HiSeq4000 (Illumina). The quality of the raw FASTQ files were checked with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). For samples originating from mouse xenografts, FASTQ reads were classified as originating from either mouse (GRCm38) or human (GRCh38) genomes using xenome[47], and human-specific reads were used for mapping. Reads were mapped to human reference GRCh38 using STAR (v2.4.1d, 2-pass mode)[48]. Gene expression was estimated using cufflinks (v2.2.1, default parameters) and HTSeq (v0.6.1)[49,50]. Differential expression analyses were performed using DESeq2 (v1.14.1)[51]. Pre-ranked gene set enrichment analyses were performed on the DESeq2 $\log_2$ fold changes[52]. Prior to any unsupervised analyses, expression counts were transformed using variance stabilizing transformation using the DESeq2 R package. Gene signatures used in the study are listed in Supplementary Table 5. Differentially expressed gene sets and their associated statistics are listed in Supplementary Table 6. To detect potential copy number changes, positional gene enrichment analysis (PGE) was performed on the upregulated and downregulated differentially expressed genes, separately (padj ≤ 0.1)[53]. Only significant regions with four or more genes and with $P \leq 0.01$ were kept for further analysis. Circos plots were made using the circlize R package[54].

Reverse-transcriptase quantitative polymerase chain reaction. Cells were collected into trizol reagent (Thermo Fisher Scientific) and total RNA was extracted using 'Pureling RNA mini kit' (Thermo Fisher Scientific) according to the manufacturer's instructions. Total RNA (5 µg) was used for RT–PCR using the RNA to cDNA Ecodry premix (oligo dT) cDNA synthesis kit (Clontech) according to the manufacturer's instructions. Resulting cDNA corresponding to 50 ng total RNA was used in each 20 µl of quantitative real time PCR reaction. qRT–PCR was performed using SybrGreen master mix (Biorad) and the relative expression of each gene was calculated after normalizing to *ACTB* endogenous control and using the comparative ΔC_t method. A list of the primers used is in Supplementary Table 4.

Single-cell RNA sequencing. Cells were trypsinized and resuspended in PBS. Twenty-one microlitres of a cellular suspension at 400 cells per microlitre, >95% viability, were loaded onto to the 10X Genomics Chromium platform to generate barcoded single-cell GEMs. scRNA-seq libraries were prepared according to 10X Genomics specifications (Single Cell 3′ Reagent Kits User Guide PN-120233, 10X Genomics). GEM-reverse transcription (RT) (55 °C for 2 h, 85 °C for 5 min; held at 4 °C) was performed in a C1000 Touch Thermal cycler with 96-Deep Well Reaction Module (Bio-Rad). After RT, GEMs were broken and the single-strand cDNA was cleaned up with DynaBeads MyOne Silane Beads (Thermo Fisher Scientific) and SPRIselect Reagent Kit (0.6 × SPRI; Beckman Coulter). cDNA was amplified for 14 cycles using the C1000 Touch Thermal cycler with 96-Deep Well Reaction Module (98 °C for 3 min; 98 °C for 15 s, 67 °C for 20 s, and 72 °C for 1 min × 14 cycles; 72 °C for 1 min; held at 4 °C). The quality of the cDNA was analysed using an Agilent Bioanalyzer 2100. The resulting cDNA was sheared to ~200 bp using a Covaris S220 instrument (Covaris) and cleaned using 0.6× SPRI beads. The products were end-repaired, 'A'-tailed and ligated to adaptors provided in the kit. A unique sample index for each library was introduced through 10 cycles of PCR amplification using the indexes provided in the kit (98 °C for 45 s; 98 °C for 20 s, 60 °C for 30 s, and 72 °C for 20 s × 14 cycles; 72 °C for 1 min; held at 4 °C). After two SPRI cleanups, libraries were quantified using Qubit fluorometric quantification (Thermo Fisher Scientific) and the quality assessed on an Agilent Bioanalyzer 2100. Four libraries were pooled and clustered on a HiSeq2500 in rapid mode at 10 pM on a pair end read flow cell and sequenced for 98 cycles of R1, followed by 14 bp I7 Index (10X Barcode), 8 bp I5 Index (sample index) and 10 bp on R2 (UMI). Primary processing of sequencing images was done using Illumina's Real Time Analysis software. Demultiplexing and post processing was done using the 10X Genomics Cell Ranger pipeline per the manufacturer's recommendations. scRNA-seq data were processed from raw reads to a molecule count array using the Cell Ranger pipeline[55]. Additionally, to minimize the effects of experimental artefacts on the analysis, data were pre-processed to filter out cells with low total molecule counts (library size), low complexity and high mitochondrial content, identified by a bimodal fit. The remaining cells were normalized by dividing the expression level of each gene in a cell by its total library size and then scaling by the median library size of all cells. After normalizing by library size, we performed principal component analysis to improve the robustness of the constructed Markov matrix generated when computing diffusion eigenvalues for imputation of dropout noise[56]. We chose the number of principal components to retain approximately 80% of variance in the data and excluded the first principal component, which was highly correlated with library size. Imputation of both the normalized and unnormalized count matrix was performed using a Markov matrix raised to the power of 3 (power corresponds the approximate number of weighted nearest neighbours) and with a gene expression distribution computed according to 21 nearest neighbouring cells, as described[56]. Our analysis was robust to imputation and we obtained similar results without imputed data (not shown). Subpopulations were identified using Phenograph[57] and genes differentially expressed in at least one subpopulation were identified by the Kruskal–Wallis rank statistic using a bootstrapping method for random downsampling of matched molecule and cell counts from each subpopulation. t-SNE was used to visualize subpopulation structure based on the first 20 principle components of the imputed count matrix, subsetted by the top

5,150 differentially expressed genes (FDR q of Kruskal–Wallis rank statistic <0.05). The mean expression of key gene signatures in population M versus other subpopulations was z-normalized and visualized using violin plots. All gene signatures are annotated in Supplementary Table 5. The correlation between gene signatures was computed using the Spearman rank correlation coefficient according to the mean expression of all genes per signature per cell. Ward's minimum variance method was applied to hierarchically cluster cells by their normalized expression of differentially expressed EMT genes.

Patient survival analysis. Genes used for survival analysis are listed in Supplementary Table 5. For the meta-analysis cohort, we used aggregate data from KMPlot[16,58] (http://www.kmplot.com) using only JetSet best probe set and auto-selection for best cutoff between the 25th and 75th percentiles. For the validation cohort in which DMFS data were available[17], we used the z-normalized expression data for a data set and the median value was used as a cutoff. DMFS curves were compared using the log-rank test.

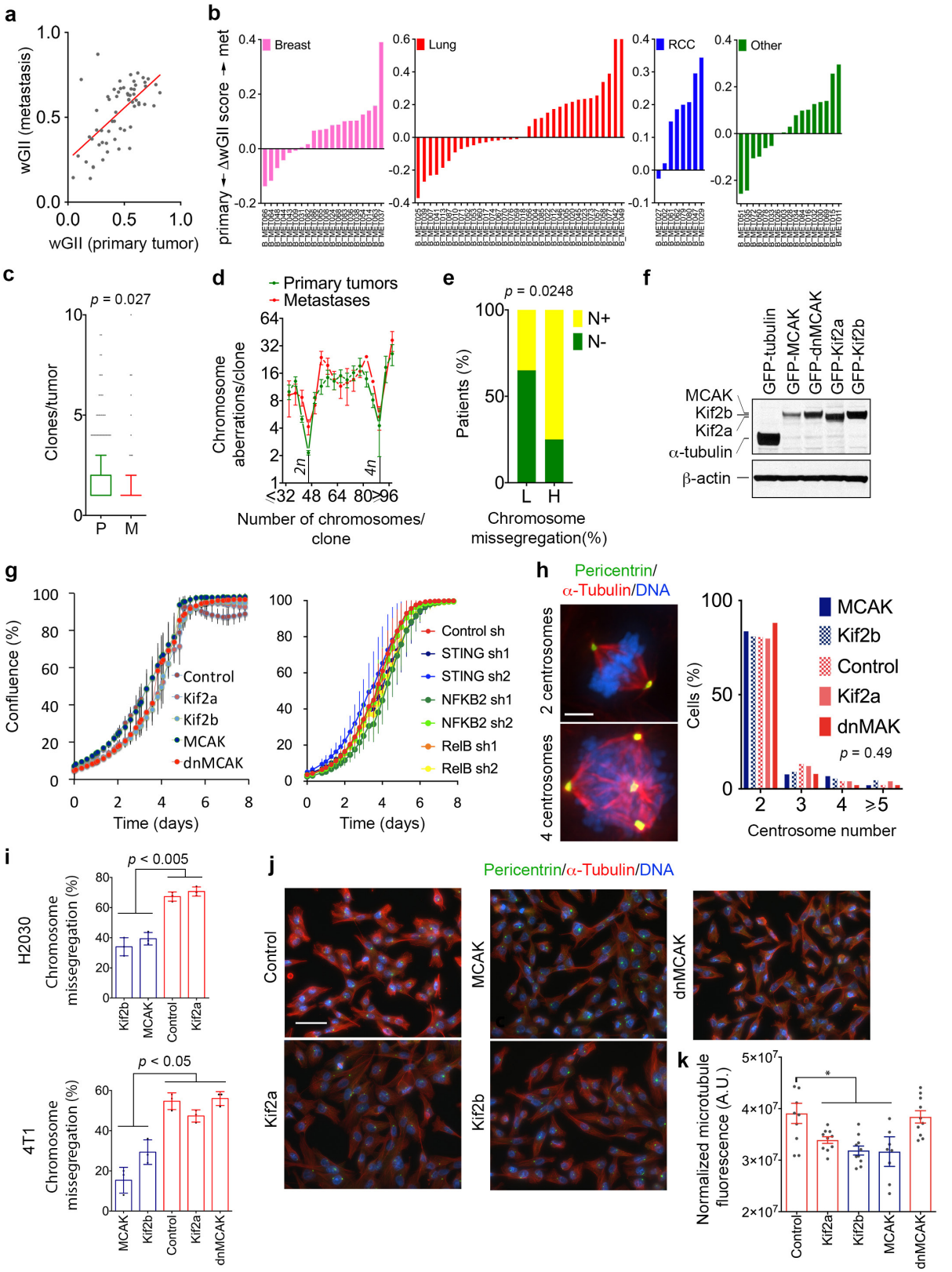
***In vitro* invasion and migration assays.** For the invasion and migration/chemotaxis assays we used the CytoSelect cell invasion (CBA-110) and cell migration (CBA-100) kits, respectively. In brief, 3×10^5 cells were suspended in serum-free medium and placed on top of the membrane. Medium containing serum was placed at the bottom and cells that had invaded to the inferior surface of the collagen membrane were stained and counted 18–24 h later. For experiments involving transient transfection, cells were transfected, and thymidine (2 mM) was added 18 h later. Cells were plated on the membrane 3 days after transfection. For the chemotaxis assay, we used a colorimetric approach (OD 560 nm) for quantification. For the scratch assay, cells were treated with mitomycin C (10 μg ml^{-1}) for 1 h when they reached >90% confluence and then placed in DMEM containing 1% FBS. Wounds were applied using a p200 pipette tip and images of the wounds were taken immediately and at subsequent regular intervals. ImageJ was used for quantification of wound surface area.

Quantification of cytosolic DNA. Approximately 1×10^7 cells were lysed and the nuclear, cytosolic, and mitochondrial fractions were obtained using the mitochondrial isolation kit (Thermo Fisher Scientific, 89874). Protease inhibitors were not used to enable subsequent DNA purification. Mitochondria were purified by centrifugation at 12,000g to minimize their contamination in the cytosolic fraction. DNA was subsequently isolated from the nuclear, cytosolic, mitochondrial fractions using the Qiagen DNeasy blood and tissue kit (Qiagen, 69506) and dsDNA was quantified using Qubit 2.0 (Invitrogen) with Qubit dsDNA HS Reagent.

Code availability. All custom code, statistical analysis, and visualizations were performed in Python or R, and used Nextflow to manage some of the computational pipelines[59]. Code for the RNA sequencing analysis is available online at: https://github.com/murphycj/manuscripts/tree/master/BakhoumEtAl2017. The live-cell tracking MATLAB 2016a code can be found at https://github.com/Lammerding/MATLAB-CellTracking.

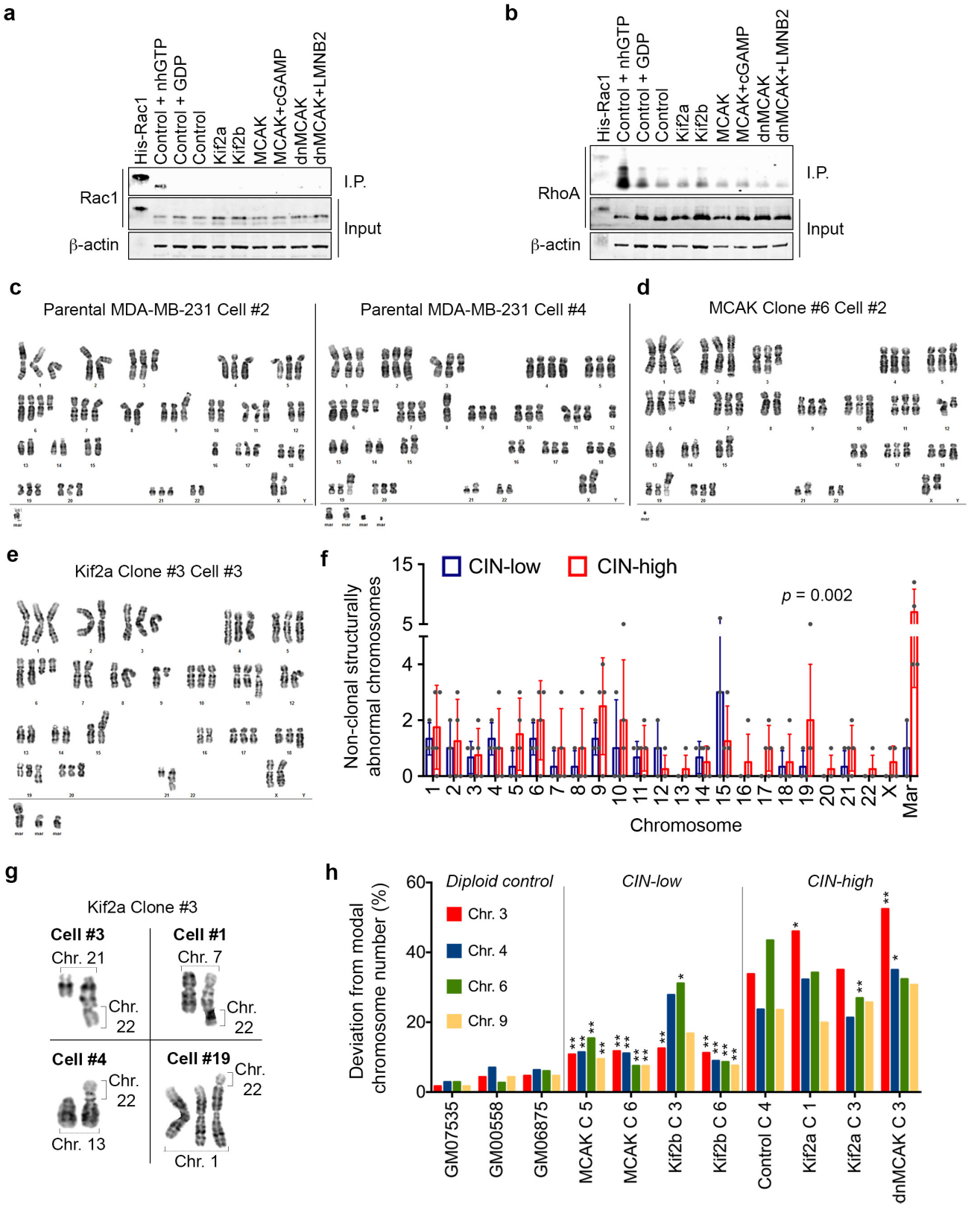
Data availability. Source data for Figs 1–3, 5, 6, and Extended Data Figs 1–3, 5–10 are provided with the paper. Single-cell RNA sequencing data (shown in Figs 4, 5 and Extended Data Fig. 6) have been deposited in the Sequence Read Archive under accession number SRP104750. Bulk RNA-seq data (shown in Extended Data Figs 4, 5) have been deposited in the Gene Expression Omnibus under accession number GSE98183.

43. McGranahan, N. *et al.* Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* **351,** 1463–1469 (2016).
44. Bakhoum, S. F., Danilova, O. V., Kaur, P., Levy, N. B. & Compton, D. A. Chromosomal instability substantiates poor prognosis in patients with diffuse large B-cell lymphoma. *Clin. Cancer Res.* **17,** 7704–7711 (2011).
45. Zaki, B. I., Suriawinata, A. A., Eastman, A. R., Garner, K. M. & Bakhoum, S. F. Chromosomal instability portends superior response of rectal adenocarcinoma to chemoradiation therapy. *Cancer* **120,** 1733–1742 (2014).
46. Davidson, P. M., Denais, C., Bakshi, M. C. & Lammerding, J. Nuclear deformability constitutes a rate-limiting step during cell migration in 3-D environments. *Cell. Mol. Bioeng.* **7,** 293–306 (2014).
47. Conway, T. *et al.* Xenome—a tool for classifying reads from xenograft samples. *Bioinformatics* **28,** i172–i178 (2012).
48. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
49. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28,** 511–515 (2010).
50. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31,** 166–169 (2015).
51. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).
52. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102,** 15545–15550 (2005).
53. De Preter, K., Barriot, R., Speleman, F., Vandesompele, J. & Moreau, Y. Positional gene enrichment analysis of gene sets for high-resolution identification of overrepresented chromosomal regions. *Nucleic Acids Res.* **36,** e43 (2008).
54. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize implements and enhances circular visualization in R. *Bioinformatics* **30,** 2811–2812 (2014).
55. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8,** 14049 (2017).
56. van Dijk, D. *et al.* MAGIC: A diffusion-based imputation method reveals gene–gene interactions in single-cell RNA-sequencing data. Preprint at https://www.biorxiv.org/content/early/2017/02/25/111591 (2017).
57. Levine, J. H. *et al.* Data-driven phenotypic dissection of aml reveals progenitor-like cells that correlate with prognosis. *Cell* **162,** 184–197 (2015).
58. Győrffy, B., Surowiak, P., Budczies, J. & Lánczky, A. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS ONE* **8,** e82241 (2013).
59. Di Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* **35,** 316–319 (2017).



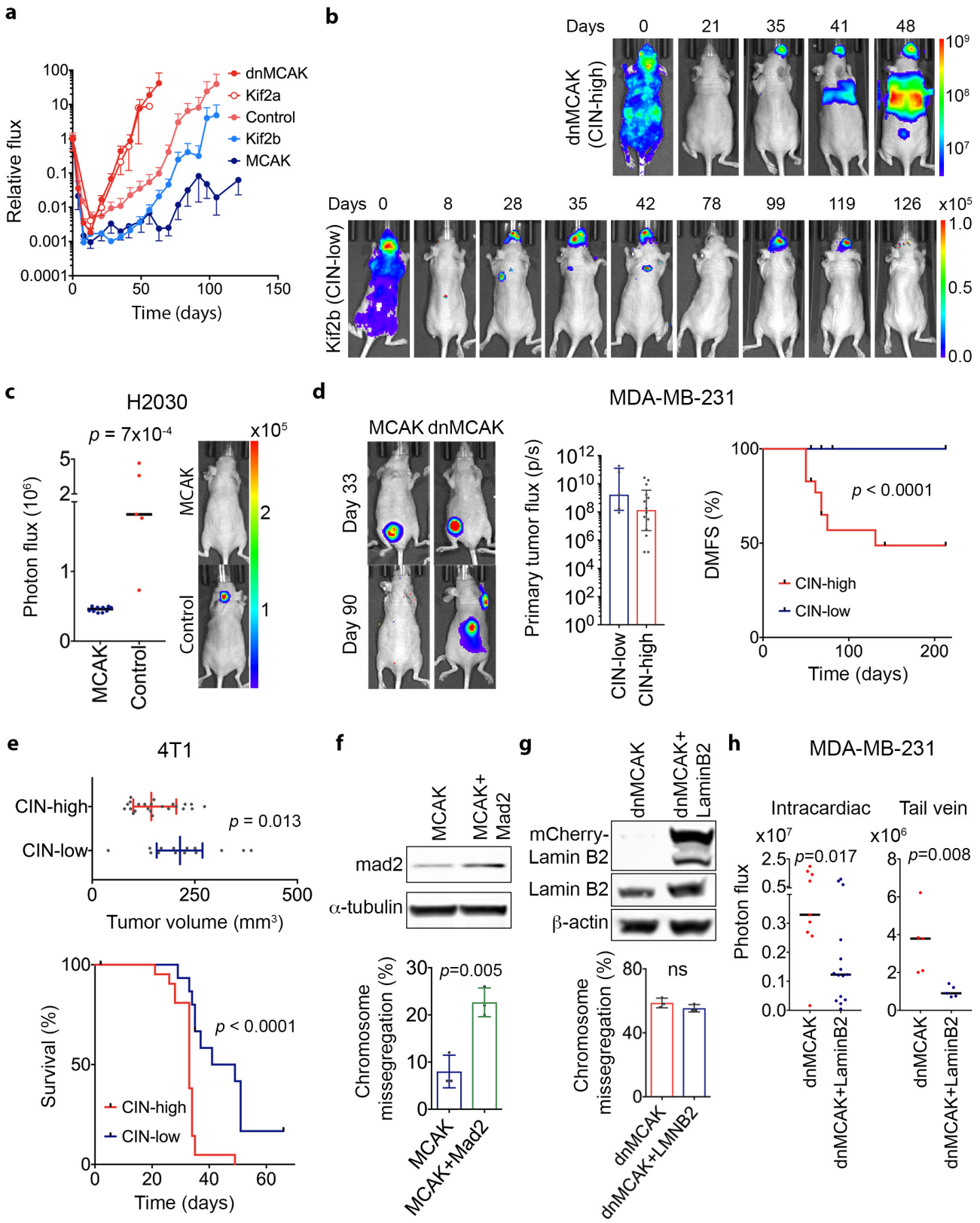
Extended Data Figure 1 | See next page for caption.

Extended Data Figure 1 | Generation of isogenic tumour models of CIN. a, wGII of brain metastases as a function of the wGII of the matched primary tumour. Red line represents linear regression, $n = 79$ patients. **b**, Differences in wGII between metastases and matched primary tumours. RCC, renal cell carcinoma; other includes melanoma, sarcoma, and ovarian, thyroid, and salivary gland cancers. **c**, Number of clones (based on single-cell karyotypes) in primary breast tumours (P; $n = 637$) or metastases (M; $n = 131$) found in the Mitelman database. Boxes represent median ± interquartile range and bars span the 10th and 90th percentiles; significance tested using two-sided Mann–Whitney test. **d**, The number of chromosome aberrations per clone as a function of the total number of chromosomes in a given clone in samples derived from primary breast tumour clones ($n = 983$) and metastatic clones ($n = 186$); data shown as mean ± s.d. **e**, Percentage of N− or N+ patients as a function of chromosome missegregation frequency ($n = 20$ patients per condition); significance tested using two-sided Fisher's exact test. **f**, Immunoblots of cells expressing various GFP-tagged kinesin-13 proteins stained using anti-GFP antibody; β-actin used as a loading control, two independent experiments performed. **g**, Cellular confluence as a function of time in MDA-MB-231 cells expressing various kinesin-13 proteins or dnMCAK-expressing cells depleted of components of the cytosolic DNA-sensing machinery or the noncanonical NF-κB pathway. Data shown as mean ± s.d., $n = 4$ independent experiments. **h**, Left, MCAK- and dnMCAK-expressing cells stained for microtubules (α-tubulin, DM1A), centrosomes (pericentrin) and DNA (DAPI). Scale bar, 5 μm; two independent experiments performed. Right, frequency distribution of the number of pericentrin foci per cell. Significance tested using ANOVA. $n = 100$ cells per condition, two independent experiments performed. **i**, Chromosome missegregation in H2030 and 4T1 cells expressing kinesin-13 proteins. Data shown as mean ± s.d., $n = 150$ cells, three independent experiments performed, significance tested using two-sided *t*-test. **j**, Cells expressing kinesin-13 proteins stained for microtubules (DM1A), centrosomes (pericentrin) and DNA (DAPI). Scale bar, 50 μm, two independent experiments performed. **k**, Fluorescence normalized to cell count of MDA-MB0-231 cells expressing kinesin-13 proteins. Data shown as mean ± s.e.m., $^*P < 0.05$, two-sided *t*-test, $n = 10$ high-power fields encompassing 477–612 cells, two independent experiments performed.



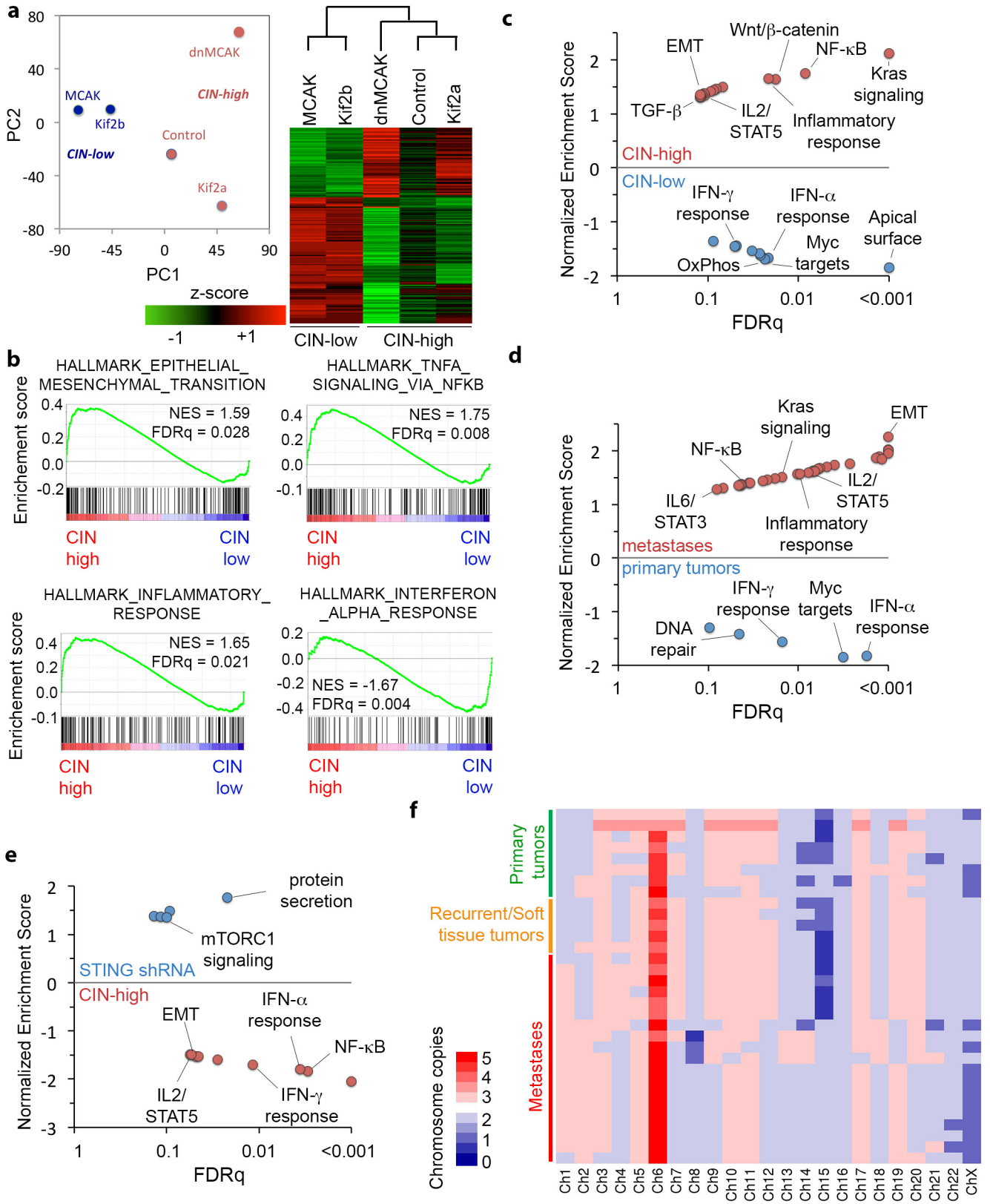
Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | Karyotype analyses of human tumour cells. **a**, **b**, Immunoblots showing total RAC1 (**a**) or RHOA (**b**) levels as well as RAC1 or RHOA that was pulled down using antibodies that were specific to the GTP-bound form of RAC1 (**a**) or RHOA (**b**). Positive and negative controls were total MDA-MB-231 cell lysates supplanted with non-hydrolysable GTP (nhGTP) and GDP, respectively. β-actin was used as a loading control, two independent experiments performed. **c**–**e**, Representative karyotypes (DAPI-banding) from parental MDA-MB-231 cells (**c**) or populations derived from single cells expressing MCAK (**d**) or KIF2A (**e**) that were allowed to divide for 30 days. **f**, The number of non-clonal (present in less than 25% of the cells in a single clone) structurally abnormal chromosomes in CIN-low or CIN- high MDA-MB-231 cells. Mar, chromosomes so structurally abnormal that they could not be identified by conventional banding; data shown as mean ± s.d., $n = 140$ cells from 7 clonal populations, significance tested using two-way ANOVA test. **g**, Examples taken from four distinct cells belonging to the same clonal population (derived from a single KIF2A-expressing cell) showing convergent translocations involving chromosome 22 with four other chromosomes. **h**, Deviation from modal chromosome number in single-cell-derived clones grown for 30 days. Four chromosomes were assayed for each clone using centromere-specific probes. $*P < 0.05$, $**P < 0.005$ compared to control clone 4 by two-sided χ^2-test, $n = 300$ cells per clone. Diploid controls were used to determine the false-positive rate of the centromeric probes.

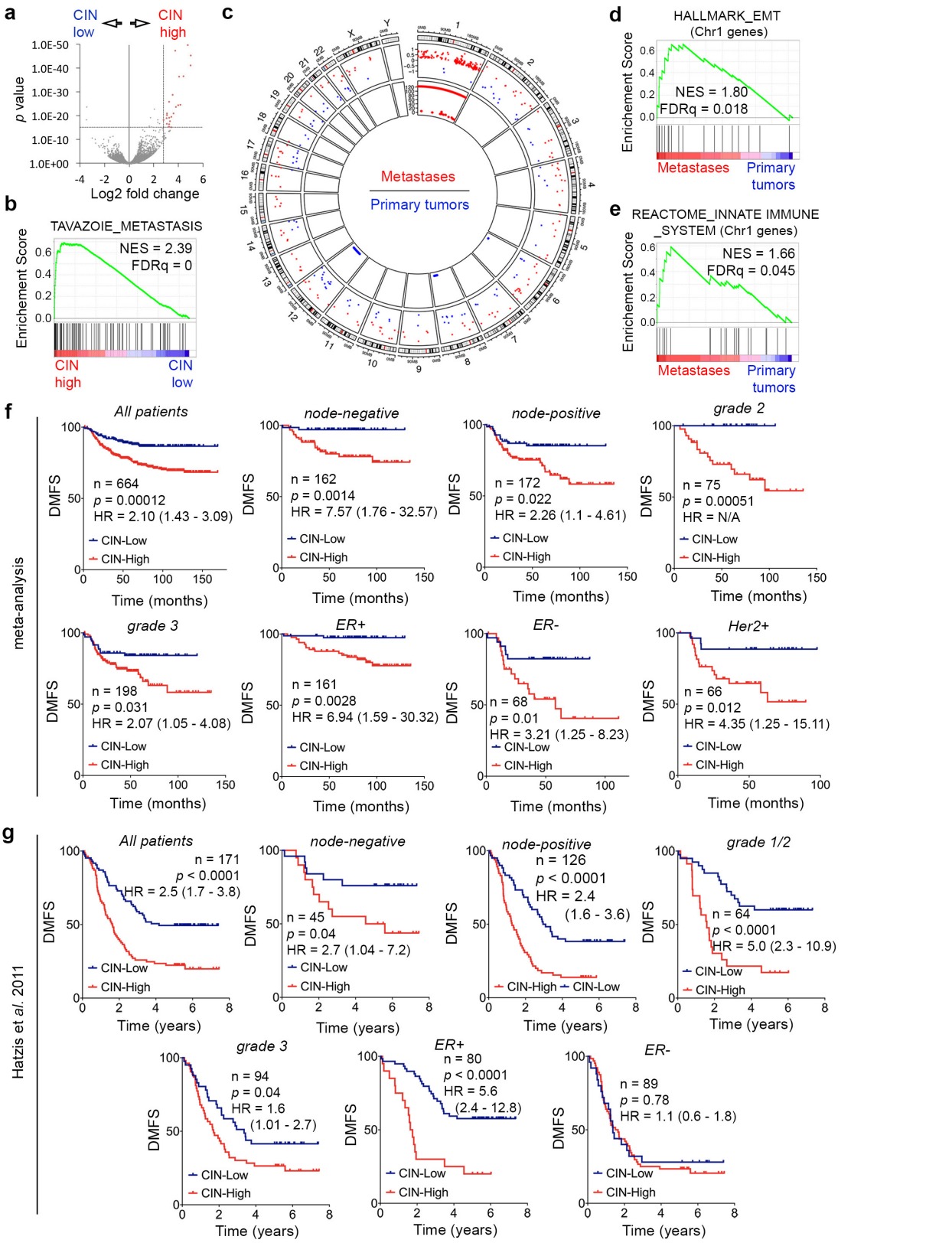


Extended Data Figure 3 | See next page for caption.

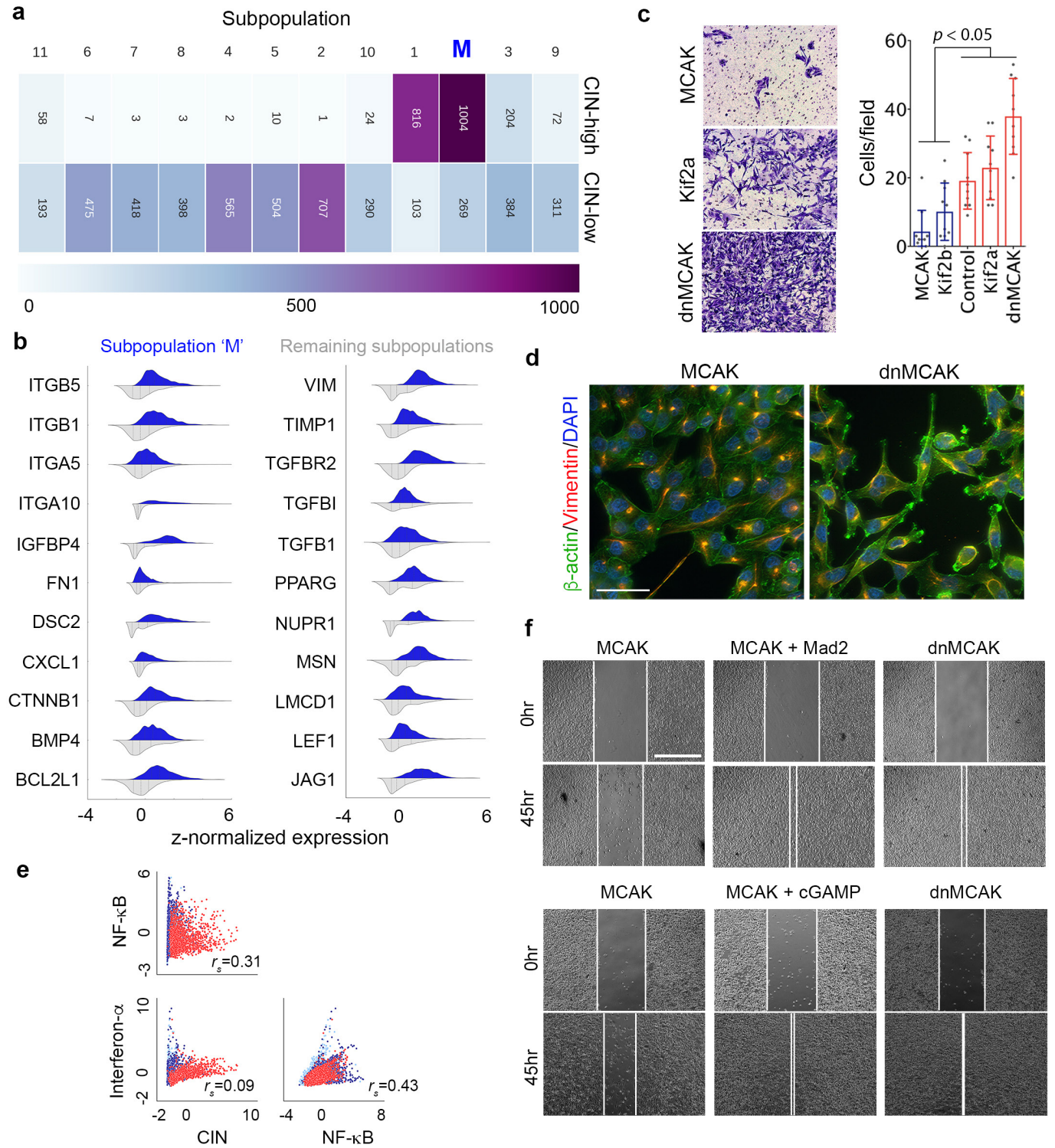
Extended Data Figure 3 | CIN promotes the formation and maintenance of metastasis. a, Normalized photon flux over time of whole animals injected with MDA-MB-231 cells expressing kinesin-13 proteins. Data shown as mean ± s.e.m. $n = 8$ (MCAK), 7 (KIF2B), 5 (control), 4 (KIF2A), and 9 (dnMCAK) mice per group; three independent experiments performed. **b**, Representative images of mice injected with MDA-MB-231 cells expressing dnMCAK (above) or KIF2B (below) with disease burden tracked using BLI; three independent experiments performed. **c**, Photon flux (p s^{-1}) of whole animals imaged 5 weeks after intracardiac injection with control or MCAK-expressing H2030 cells. Horizontal bars represent the mean, significance tested using two-sided Mann–Whitney test, $n = 10$ mice in the MCAK group and 5 mice in the control group. **d**, Left, representative BLI images (from two independent experiments) of mice orthotopically transplanted with MDA-MB-231 cells. Images taken before (day 33) and after (day 90) tumour excision. Metastasis can be detected in the mouse transplanted with dnMCAK-expressing cells at day 90. Middle, total flux (p s^{-1}) emitted from primary tumours 52 days after transplantation. Data shown as mean ± s.d., $n = 5$ (CIN-low) and 14 (CIN-high) mice, $P = 0.13$, two-sided Mann–Whitney test. Right, DMFS of mice orthotopically transplanted with MDA-MB-231 cells with various levels of CIN. $n = 15$ (CIN-low) and 29 (CIN-high) mice, pairwise significance tested with two-sided log-rank test. **e**, Tumour volume at 8 days (top) and survival (bottom) of mice transplanted with mouse 4T1 cells into the mammary fat pad. Bars represent median ± interquartile range, pairwise significance tested with two-sided *t*-test (top) and two-sided log-rank test (bottom). $n = 20$ (CIN-low) and 30 (CIN-high) mice. **f**, Top, immunoblots of MDA-MB-231 cells overexpressing MCAK or MCAK and MAD2 stained for MAD2 using anti-MAD2 antibody with α-tubulin used as a loading control; three independent experiments performed. Bottom, percentage of anaphase cells exhibiting evidence of chromosome missegregation in cells overexpressing MCAK or MCAK and MAD2. Data shown as mean ± s.d., $n = 150$ cells, three experiments performed, significance tested using two-sided *t*-test. **g**, Top, immunoblots of MDA-MB-231 cells overexpressing dnMCAK or dnMCAK and lamin B2 stained for lamin B2 using anti-lamin B2 antibody with β-actin used as a loading control. Two experiments performed. Bottom, percentage of anaphase cells exhibiting evidence of chromosome missegregation in cells overexpressing dnMCAK or dnMCAK and lamin B2. Data shown as mean ± s.d., $n = 150$ cells, three experiments performed, significance tested using two-sided *t*-test. **h**, Photon flux (p s^{-1}) of whole animals after intracardiac (left) or tail vein (right) injection with MDA-MB-231 cells expressing dnMCAK or dnMCAK and lamin B2. Bars represent the median, significance tested using two-sided Mann–Whitney test, $n = 9$ (dnMCAK), 15 (dnMCAK and Lamin B2) mice in the intracardiac injection cohort and 5 mice per group in the tail vein injection cohort.



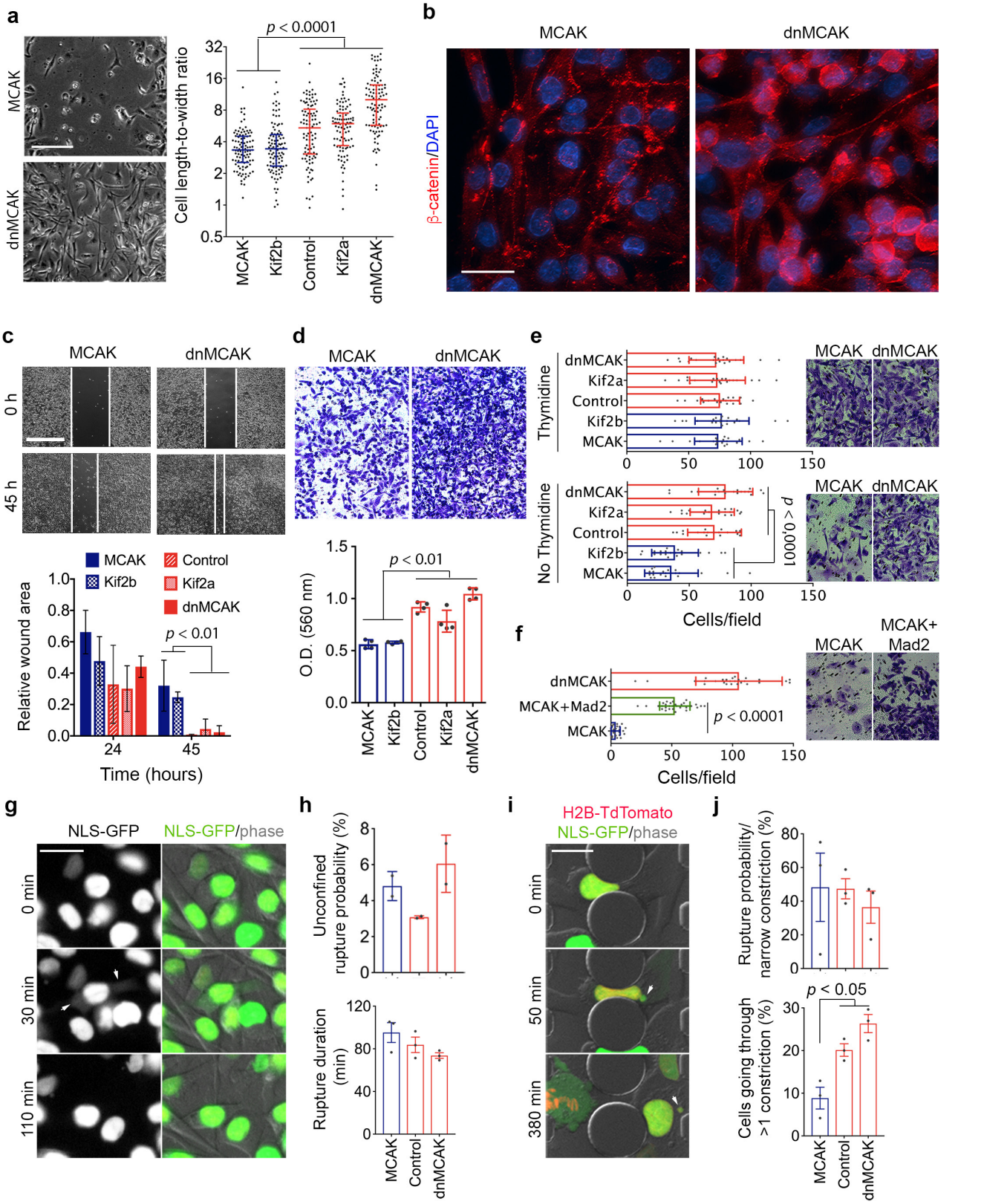
Extended Data Figure 4 | Transcriptional consequences of CIN in cancer cells. a, **b**, Principal component analysis (left) and unsupervised clustering (right) of five MDA-MB-231 cell lines expressing different kinesin-13 proteins based on bulk RNA expression data. **b**–**e**, Gene set enrichment analysis results showing HALLMARK gene sets that are highly enriched in CIN-high (control, KIF2A, and dnMCAK) compared with CIN-low cells (MCAK and KIF2B) (**b**, **c**) or STING-depleted cells (**e**), or after comparing metastases with primary tumours (**d**). Significance tested using one-sided weighted Smirnov–Kolmogorov test corrected for multiple tests. **f**, Heat map of consensus chromosomal karyotypes of cells derived from primary tumours and metastases showing selective increase in chromosome 1 copy number in metastases compared with primary tumours.



Extended Data Figure 5 | Prognostic impact of CIN signature. **a**, Volcano plot showing genes that were differentially expressed between CIN-high and CIN-low MDA-MB-231 cells. Red data points denote genes subsequently used for determining the CIN signature. **b**–**e**, Enrichment plots for all differentially expressed genes (**a**) or those on chromosome 1 (**d**, **e**). Circos plot (**c**) shows genomic location (outer circle), $\log_2$ fold expression of genes significantly differentially expressed in metastases compared to primary tumours (middle circles), and $\log_{10} P$ (inner circle) for genomic amplifications (red) or deletions (blue) in metastases relative to primary tumours. $n = 2$ (CIN-low), 3 (CIN-high), 11 (primary tumours), 28 (metastases). Significance tested using two-sided Wald test (**a**), one-sided weighted Smirnov–Kolmogorov test (**b**, **d**, **e**), and one-sided hypergeometric test (**c**), all corrected for multiple testing. **f**, **g**, DMFS of breast cancer patients stratified by lymph node status, grade, and receptor status, from a meta-analysis (**f**, $n = 664$ patients) or a validation cohort (**g**, $n = 171$ patients) divided on the basis of average expression of the CIN gene expression. Significance tested using two-sided log-rank test.

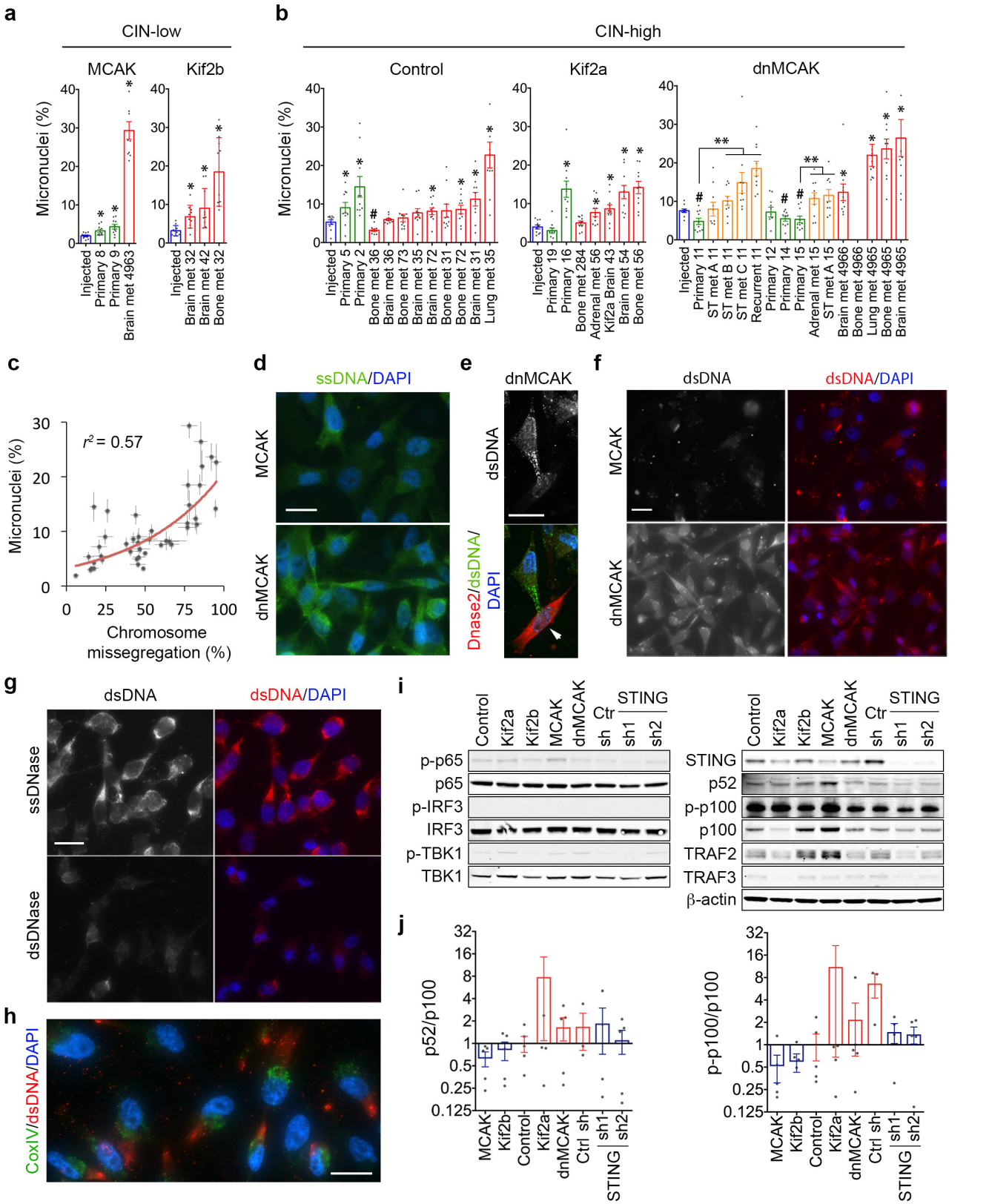


Extended Data Figure 6 | Single-cell sequencing and population detection. a, The cellular composition of every subpopulation presented Fig. 4b. **b**, Violin plots showing expression probability density of key metastasis and invasion genes in a subpopulation of cells ($n = 1,273$ cells) enriched for EMT and CIN genes (subpopulation M) compared with the remaining subpopulations ($n = 5,548$ cells) that were identified using graph-based unsupervised *K*-nearest neighbour embedding. **c**, Representative low-power field images (left) and numbers (right) of MDA-MB-231 cells that invaded through a collagen membrane within 18 h of culture. Data shown as mean ± s.d., significance tested using two-sided Mann–Whitney test, $n = 10$ high-power fields, two independent experiments performed. **d**, Representative images of MDA-MB-231 cells expressing MCAK or dnMCAK stained for β-actin, vimentin, and DNA. Scale bar, 50 μm, $n = 2$ independent experiments. **e**, Single-cell correlation plots between CIN signature genes, canonical NF-κB and type I interferon target genes, $n = 6,821$ cells. **e**, Representative phase-contrast images of a wound-healing assay of MDA-MB-231 cells expressing MCAK, MCAK and MAD2 or dnMCAK, and MCAK-expressing cells treated with cGAMP. Scale bar, 800 μm, four experiments performed.



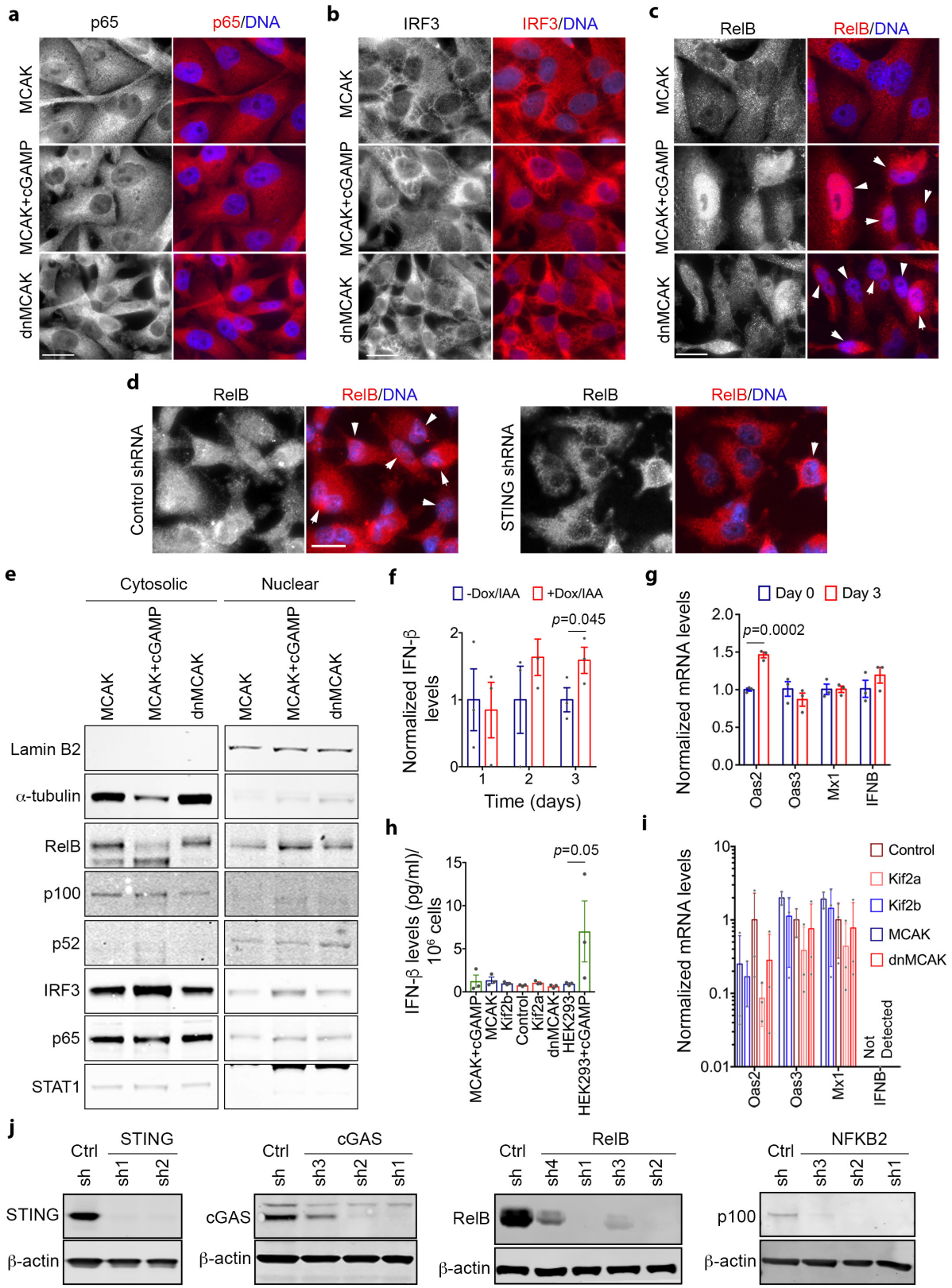
Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | CIN promotes *in vitro* invasion and migration. a, Left, representative phase-contrast images of MDA-MB-231 cells in the wound area, 36 h after wound creation. Four experiments performed. Right, length-to-width ratio of cells expressing different kinesin-13 proteins. Bars span the interquartile range, $n = 100$ cells, two independent experiments performed, significance tested using two-sided Mann–Whitney test. **b**, Representative MDA-MB-231 cells stained for β-catenin (anti-β-catenin antibody) or DNA (DAPI). Changes in β-catenin are seen upon alteration of CIN; it is enriched at cell–cell junctions in MCAK-expressing cells but is found in the cytoplasm and nucleus in dnMCAK-expressing cells. Scale bar, 30 μm, two experiments performed. **c**, Top, phase-contrast images of a wound-healing assay of cells expressing kinesin-13 proteins. Scale bar, 800 μm, two experiments performed. Bottom, wound area (normalized to the 0 h time point) 24 h and 45 h after wound creation. Data shown as mean ± s.d., $n = 4$ experiments, significance tested using two-sided *t*-test. **d**, Top, low-power field images of MDA-MB-231 cells that have migrated through a polycarbonate membrane containing 8-μm pores within 18 h of culture. Bottom, normalized OD of cells scraped from the bottom of the membrane. Data shown as mean ± s.e.m., significance tested using two-sided *t*-test, $n = 3$ experiments. **e**, **f**, Left, number of MDA-MB-231 cells that have successfully invaded through a collagen basement membrane 24 h after plating. Data shown as mean ± s.d., $n = 20$ high power fields from two independent experiments, significance tested using two-sided Mann–Whitney test. Right, representative images from high-power fields. Two independent experiments performed. **g**, **i**, Representative time-lapse fluorescence and phase-contrast image sequences of control cells expressing NLS–GFP undergoing unconfined migration (**g**) or going through $1 \times 5\text{-}\mu m^2$ constrictions (**i**). Scale bars, 20 μm. Arrows in **g** indicate cytoplasmic NLS–GFP. Arrows in **i** indicate formation of nuclear protrusion and subsequent fragments during confined migration. Three independent experiments performed. **h**, **j**, Top, the probability of primary nuclear rupture during unconfined conditions (**h**) or after migration through $1 \times 5\text{-}\mu m^2$ constrictions (**j**). Bottom, the number of cells migrating through more than one 1-μm-wide constrictions (**j**) and the duration of nuclear rupture (**h**), as measured by the length of time for which NLS–GFP signal is observed in the cytosol. Data shown as mean ± s.e.m., $n = 3$ independent experiments (except for unconfined rupture probability, 2 independent experiments) encompassing 390–665 (**h**) and 150–336 (**j**) cells observed during unconfined and confined migration, respectively. Significance tested using two-sided *t*-test.



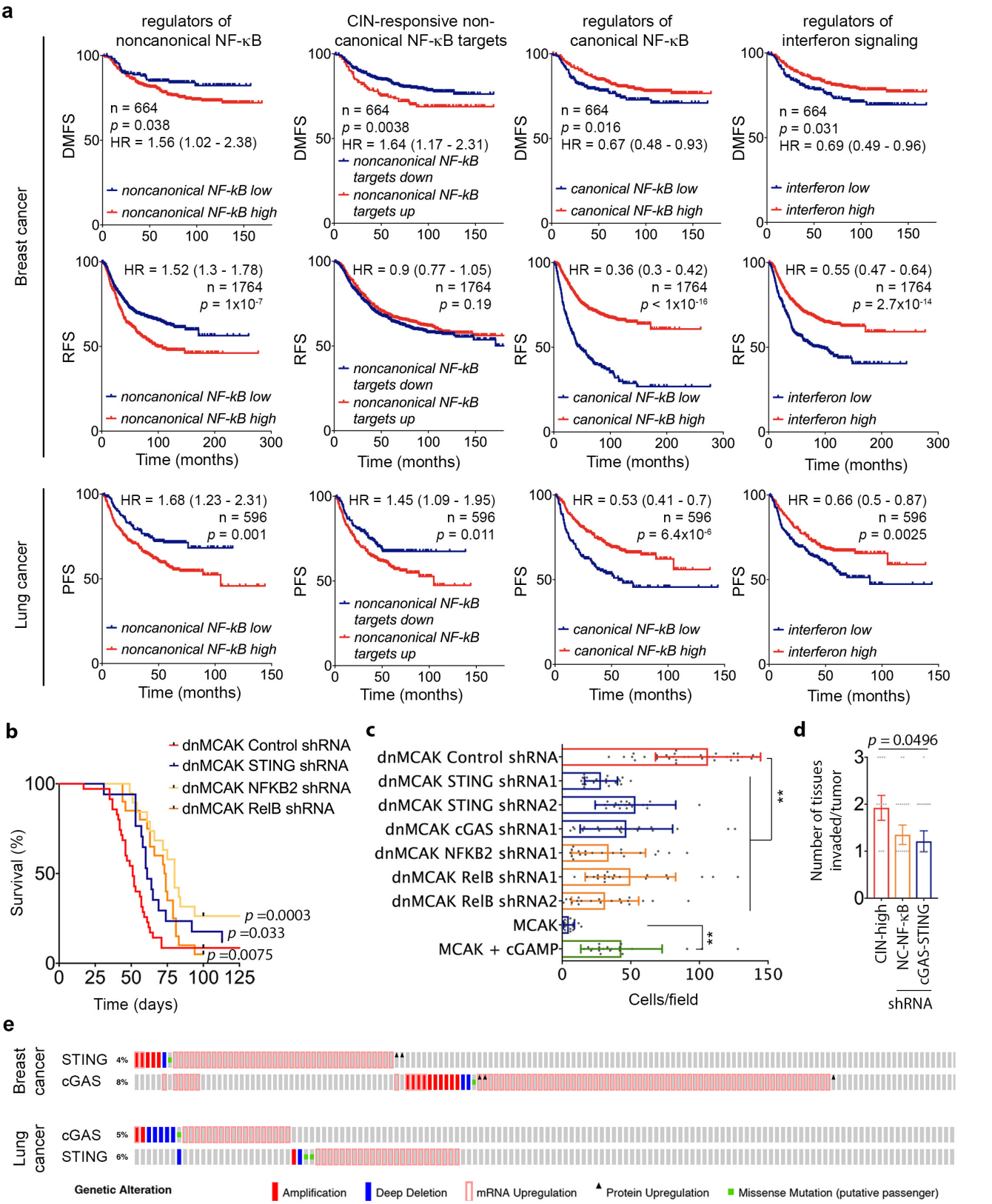
Extended Data Figure 8 | See next page for caption.

Extended Data Figure 8 | CIN generates micronuclei and cytosolic DNA. a, **b**, Percentage of micronuclei in samples depicted in Fig. 3c, d: injected cells (blue), first-passage cells derived from primary tumours (green), or metastases (orange denotes spontaneous metastases arising from primary tumours, red denotes metastases obtained from direct intracardiac implantation). Data shown as mean ± s.e.m., $n = 10$ high-power fields encompassing 500–1,500 cells per sample, three independent experiments performed, $^{*}P < 0.05$ (denotes samples with higher missegregation rates than the injected lines), $^{\#}P < 0.05$ (denotes samples with lower missegregation rates than the injected lines), $^{**}P < 0.05$ (denotes significant differences between metastases and matched primary tumours from the same animals), two-tailed *t*-test. **c**, Correlation between the percentage of cells exhibiting evidence of chromosome missegregation and the percentage of micronuclei in all injected cell lines as well as cells derived from primary tumours and metastases. Data shown as mean ± s.e.m., $n = 44$ samples. **d**–**f**, Representative images of cells stained for DNA (DAPI), cytosolic single-stranded DNA (ssDNA) (**d**), DNASE2 (RFP reporter) (**e**), or cytosolic dsDNA (**f**). Scale bar, 20 μμm, arrows in **e** denote DNASE2-expressing cell, two independent experiments performed. **g**, Representative images of dnMCAK-expressing cells treated with ssDNASE or dsDNASE for 10 min after selective plasma membrane permeabilization (using 0.02% saponin) and stained for DNA (DAPI) and cytosolic dsDNA. Scale bar, 20 μm, one experiment performed. **h**, Representative images of dnMCAK-expressing cells stained for mitochondria (anti-CoxIV antibody), DNA (DAPI) or for cytosolic DNA (anti-dsDNA antibody). Scale bar, 20 μm, two independent experiments performed. **i**, Immunoblots of lysates from cells expressing different kinesin-13 proteins, control or STING shRNA. β-actin used as a loading control. **j**, Normalized ratio of phosphorylated p52 to p100 (left) and p100 to total p100 (right) protein levels. Data shown as mean ± s.e.m., $n = 5$ independent experiments.



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Alternative response to cytosolic DNA in cancer cells. a–d, Representative images of MDA-MB-231 cells stained for DNA (DAPI) and for p65 (**a**), IRF3 (**b**), or RELB (**c**, **d**). Images were individually contrast-enhanced to emphasize nuclear versus cytosolic localization of p65, IRF3, and RELB. For quantitative comparisons of identical images, see Supplementary Fig. 3. Arrows (**c**, **d**) point to RELB-positive nuclei. Scale bars, 20 μm, three independent experiments performed. **e**, Immunoblots of fractionated lysates. α-tubulin and lamin B2 were used as loading controls for the cytoplasmic and nuclear fractions, respectively; three independent experiments performed. **f**, **h**, Interferon-β levels in conditioned medium from DLD-1 cells (**f**), MDA-MB-231 or HEK293 cells with and without cGAMP addition (**h**). Data shown as mean ± s.e.m. $n = 3$ experiments, significance tested using one-sided Mann–Whitney test. **g**, **i**, Relative levels of interferon-responsive genes obtained by RT–qPCR in DLD-1 cells (**g**) normalized to untreated conditions or MDA-MB-231 cells (**i**) normalized to control cells. Data shown as mean ± s.d. $n = 3$ experiments, significance tested using two-sided *t*-test. **j**, Immunoblots of lysates of dnMCAK-expressing cells that also co-expressed control shRNA or shRNAs targeting components of the cytosolic DNA-sensing or noncanonical NF-κB pathways. shRNA hairpins are numbered in ascending order according to the efficiency of protein knockdown. Two independent experiments performed.



Extended Data Figure 10 | See next page for caption.

Extended Data Figure 10 | Effect of cytosolic DNA-sensing pathways on prognosis. a, Distant metastasis-free survival (DMFS), relapse-free survival (RFS) and progression-free survival (PFS) of patients with breast and lung, stratified according to their expression of NF-κB and interferon pathways. Significance tested using two-sided log-rank test. **b**, Disease-specific survival of mice injected with dnMCAK-expressing MDA-MB-231 cells co-expressing control shRNA, STING shRNA, NFKB2 shRNA, or RELB shRNA. $n = 35$, 16, 19, and 20 mice in the control, STING shRNA, NFKB2 shRNA, and RELB shRNA groups, respectively; significance tested using two-sided log-rank test. **c**, Number of MDA-MB-231 cells expressing shRNA targeting genes belonging to the DNA-sensing or noncanonical NF-κB pathways that invaded through a collagen membrane within 24 h of culture. Data shown as mean ± s.d., ** $P < 0.0001$, two-sided Mann–Whitney test, $n = 20$ high-power fields, two independent experiments performed. **d**, Number of different normal tissues (vascular, neuronal, or soft tissue) invaded by orthotopically transplanted tumours. Data shown as mean ± s.e.m., *$P < 0.05$, two-tailed t-test, $n = 13$ tumours (CIN-high), 20 tumours (noncanonical NF-κB depleted), 19 tumours (cGAS-STING depleted). **e**, Oncoprints showing genomic alterations in STING (TMEM173) and cGAS (MB21D1) in breast and lung cancers from the TCGA database.

nature research

Corresponding author(s): Lewis Cantley, PhD

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

For disease-specific survival in MDA-MB-231 experiments, power analysis indicated that 10 mice per group will be sufficient to detect a difference at relative hazard ratios of <0.2 or >5 with 80% power and 95% confidence, given a median disease-specific survival of 3 months in the control group and a total follow up period of 250 days. For the 4T1 experiments, Power analysis indicates that 10 mice per group will be sufficient to detect a difference at relative hazard ratios of <0.25 or >4.0 with 80% power and 95% confidence, given a median survival of 58 days in the control group and a total follow up period of 180 days.

2. Data exclusions

Describe any data exclusions.

No data were excluded

3. Replication

Describe whether the experimental findings were reliably reproduced.

No attempts for replication failed. Immunofluorescence and immunoblot experiments were performed in three or more biological replicates (see figure legends). Key animal experiments were performed in up to 5 independent experiments

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

No method of randomization was used

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Investigators were not blinded to group allocation.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Bulk RNA seq code is made publicly available (link in the methods section). Single cell RNAseq was analyzed using Python and already-published data and was appropriately referenced in the methods section. Wound healing assay (surface area) was analyzed using ImageJ. Single cell analysis was performed using Python 3.0. Cell migration using microfluidics devices was performed using MATLAB 2016a. RNAseq data was performed using R

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	There are no restrictions for distribution of research materials produced during this study.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

Pre-validated antibodies were purchased from reputable sources (Cell signaling technology and abcam). Antibodies and their catalog numbers are listed in Supplementary information. Most antibodies were further validated when the target protein was depleted using shRNA.

Supplementary Table 1. Antibodies used in immunofluorescence.

Antibody against Company Catalog number
α-tubulin Sigma Aldrich T9026
β-actin Abcam ab8227
β-catenin Abcam ab16051
cGAS Sigma Aldrich HPA031700
Cox IV Abcam ab16056
dsDNA (Extended Fig. 8d) Abcam AB27156
dsDNA (Fig. 5f) Thermo Fisher Scientific MAB1293MI
Human centromere proteins Antibodies Incorporated 15-234-0001
IRF3 Abcam ab68481
p65 Abcam ab16502
Pericentrin Abcam ab4448
RelB Abcam ab180127
ssDNA Thermo Fisher Scientific MAB3299MI
Vimentin Abcam ab201637
STING Abcam ab181125

Supplementary Table 2. Antibodies used in immunoblots.

Antibody against Company Catalog number
β-actin Abcam ab8227
cGAS Sigma Aldrich HPA031700
GFP Life Technologies A11122
IRF3 Abcam ab68481
p100/p52 Cell Signaling Technology 4882
p65 Abcam ab16502
phospho-IRF3 Cell Signaling Technology 4947
phospho-p100 Abcam 194919
phospho-p65 Cell Signaling Technology 3033
phospho-TBK1 Cell Signaling Technology 5483
RelB Cell Signaling Technology 4922
STING Cell Signaling Technology 13647
TBK1 Cell Signaling Technology 3013
TRAF2 Cell Signaling Technology 4712
TRAF3 Cell Signaling Technology 4729
STAT1 Abcam ab47425

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

ATCC

b. Describe the method of cell line authentication used.

None as directly purchased from ATCC

c. Report whether the cell lines were tested for mycoplasma contamination.

Yes. Cells were tested each 4-6 months.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

None of the cell lines used are listed in the ICLAC database.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Animal experiments were performed in accordance with protocols approved by the Weill Cornell Medicine Institutional Animal Care and Use Committee. There was no need to randomize animals. Investigators were not blinded to group allocation. Intracardiac injection was performed as previously described31. Briefly, cells were trypsinized and washed with PBS and a 1x105 cells (in 100 µl of PBS) were injected into the left cardiac ventricle of female athymic 6-7 week old athymic nude (nu/nu) mice (Jackson Laboratory strain 002019). 2x105 cells were injected into the tail vein cohort of animals. Mice were then immediately injected with D-luciferin (150 mgkg-1) and subjected to bioluminescence imaging (BLI) using tan IVIS Spectrum Xenogen instrument (Caliper Life Sciences) to ensure systemic dissemination of tumor cells. Metastatic burden was measured 5-7 weeks after injection using BLI. BLI images were analyzed using Living Image Software v.2.50. Disease-specific survival endpoint was met when the mice died or met the criteria for euthanasia under the IACUC protocol and had radiographic evidence of metastatic disease. For Orthotopic tumor implantation, 2.5x105 cells in 50µl of PBS were mixed 1:1 with Matrigel (BD Biosciences) and injected into the fourth mammary fat pad. Only one tumor was implanted per animal. MDA-MB-231 primary tumors were surgically excited before they reached ~1.5 cm in the largest dimension (which was the maximum allowable under our IACUC protocol) and metastatic dissemination was assessed using BLI imaging at 1-3 week intervals for up to 30 weeks. Distant metastasis-free survival endpoint was met when BLI signal was seen outside of site of primary tumor transplantation. 4T1 tumors were excised 9 days after implantation. To derive short-term culture from primary tumors and metastases, anesthetized animals (isofluorane) were imaged then sacrificed. Ex-vivo BLI was subsequently performed on harvested organs to define the precise location of the metastatic lesion. Primary tumors and metastases were subsequently mechanically dissociated and cultured in DMEM with selection media (G418 or hygromycin) to select for tumor cells and exclude host cells. All subsequent assays (karyotyping, RNAseq, immunofluorescence, and subcellular fractionation) were performed after a single passage from the primary sample. To assess chromosome missegregation from primary tumor-derived and metastases-derived cells, we performed high-resolution immunofluorescence analysis on passage #1 cells, staining for DNA (DAPI) and centromeres (ACA). Cells with DNA or centromere staining in the middle of the anaphase plate was taken as evidence of chromosome missegregation.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

All human samples reported in the manuscript were from studies that were previously published and are adequately referenced in the manuscript. Details of patient characteristics - if available - is listed in the referenced manuscript.